Structural and functional studies of CARP and titin N2A protein complexes involved in the stress-response of muscle to apoptotic stimuli

Tiankun Zhou

Supervisor: Prof. Olga Mayans

Dr.Igor Barsukov

Institute of Integrative Biology

Structural Biology Unit

University of Liverpool

Crown Street, L69 7ZB

Liverpool, UK

December 2015

Table of contents

Abstract

The cardiac ankyrin repeat protein (CARP) has been reported to bind to the N2A region of the giant sarcomeric protein titin. The interaction allegedly plays a role in preventing cardiomyocyte apoptosis, thereby protecting against heart failure (HF). The latter is now a leading cause of mortality in the developed world. In this work, we aimed to characterize the CARP:titin-N2A interaction with view to guide future functional studies that can clarify its significance in heart disease.

First, CARP and titin N2A constructs of diverse domain composition were designed, produced in the form of expression plasmids, and their expression established in the *E. coli* bacterial system. This resulted in the availability of high-yields of soluble and highly purified protein samples that enable subsequent biophysical and structural analysis. The CARP binding site in titin-N2A was previously suggested to map to a unique, central sequence considered to be unstructured, the UN2A domain. However, using CD, thermal denaturation and NMR measurements, we have found UN2A to be a highly helical protein with a high thermal stability. Additionally, tested by gel-filtration, we now show that the high affinity interaction of CARP:titin-N2A spans both the unique UN2A and the subsequent Ig domain, Ig81. SEC-MALLS data showed that in isolation CARP is a dimer and titin-N2A a monomer, but their complex has a 1:1 stoichiometry, indicating that the breakage of the CARP dimer is required for interaction. Although all data point to a strong binding, efforts to quantitate the K_D of the interaction were not successful. Finally, all protein samples were trialed in crystallization and initial crystalline formations obtained for CARP, domain I81 and the complex CARP:UN2A-I81. A high quality X-ray diffraction data set was recorded at 2.15 Å resolution for domain I81. For UN2A, initial 2D-NMR spectra of high-quality were obtained. Taken together, this points to the promise of structurally characterizing these proteins at the atomic level.

Data here obtained allowed us to propose a new hypothesis, where the CARP interaction spanning the UN2A-I81 domains modulate titin's stiffness both by physically increasing the resistance of this region to stretch and by inhibiting titin N2A phosphorylation.

1. Introduction

1.1 CARP's structure and its role in heart failure

1.1.1 CARP is implicated in heart failure

Heart failure (HF) is a complex clinical syndrome caused by various heart diseases - such as dilated cardiomyopathy, hypertrophic cardiomyopathy and arrhythmogenic right ventricular cardiomyopathy - that leads to an insufficient blood supplement for the daily body's need. Symptoms of HF include breathless, excessive fatigue and leg swelling (National Clinical Guideline Centre, 2010). In addition, HF is associated with cardiomyocyte apoptosis (Kang *et al.,* 2000) and can, thereby, progress to critical states. HF is now a leading cause of mortality in the developed world, with a four-year death rate of ~50% (Henkel *et al.,* 2008). Given the prevalence and high death rate of HF, a search for new treatments is highly needed.

The cardiac ankyrin repeat protein (CARP, Ankrd1; UnitProtKB Q15327) is a member of the muscle ankyrin repeat protein (MARP) family, whose expression in the heart is consistently up-regulated during HF (*Table.1.1*). Thus, it is thought to increase the resistance of cardiomyocytes to apoptosis and, thereby, protect against HF (Arimura *et al.,* 2009; Bogomolovas *et al.,* 2015).

CARP is considered to be a stretch-stress response factor in the heart, as its expression can be induced in response to stretch (Kemp *et al.,* 2000). For example, a rapid and dramatic increase of CARP expression was observed in rat cardiomyocytes after passive stretching *in vitro*, indicating that this protein might be involved in the adaptive response of the sarcomere to stretch (Miller *et al.,* 2003). Additionally, CARP was observed to localize to the sarcomeric I-band during stretch overload (Bang *et al.,* 2001; Miller *et al.,* 2003) and considered to modulate muscle compliance by interacting with I-band titin at the N2A region (see section 1.2). Furthermore, the stretch overload of cardiac muscle was shown to cause apoptosis of cardiomyocytes, thereby leading to HF (Chen *et al.,* 2015). Taken together, it is conceivable that the anti-apoptotic role of CARP is the

consequence of the CARP:titin-N2A interaction, in which CARP serves as a muscle compliance modulator in response to cardiomyocyte stretch overload. (*Fig.1.1*).

CARP was also found to play a role in the resistance to several heart diseases that lead to HF. For example, CARP overexpression is associated with dilated cardiomyopathy, hypertrophic cardiomyopathy and arrhythmogenic right ventricular cardiomyopathy (Zolk *et al.,* 2002; Nagueh *et al.,* 2004; Wei *et al.,* 2009), and these diseases can also be caused by missense mutations in the CARP gene (Arimura *et al.,* 2009; Duboscq-Bidot *et al.,* 2009; Moulik *et al.,* 2009). The relationship between CARP expression and cardiomyopathy is shown in *Table.1.1*. These CARP-disease relationships support the idea that CARP's anti-apoptotic role in HF might be based on its muscle compliance modulation function (*Fig.1.1)*.

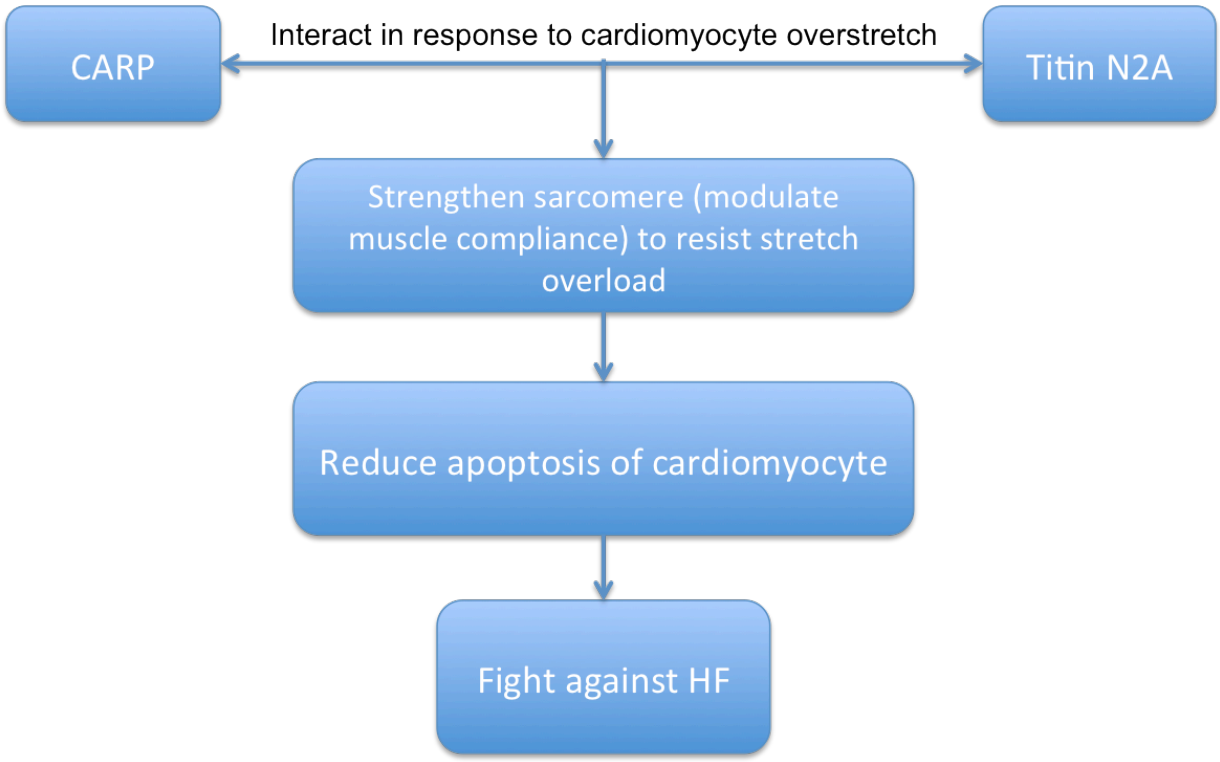


***Fig.1.1* CARP serves to resist HF by modulating sarcomere passive force**

CARP interacts with the titin N2A region to strengthen the sarcomere allowing it to resist stretch overload, thereby reducing cardiomyocyte apoptosis. This anti-apoptotic effect protects against HF.

Table.1.1 CARP expression is associated with HF (Table remade from Mikhailov *et al.,* 2008).

Model	Cardiac phenotype	Clinical phase	CARP expression	Reference
MLP-deficient mice	Early-VH	End-stage HF	Transcript (UP)	Arber *et al.,* 1997
	Late-DCM	End-stage HF	Transcript (UP)	Baumeister *et al.,* 1997
CASQ2-overexpression mice	Early-VH	Depress contractility	Transcript (UP)	Ihara et al., 2002
	Late-DCM	End-stage HF	Transcript (UP)	
Gs- α -overexpression mice	Late-DCM	Advanced HF	Transcript (UP)	Gaussin *et al.,* 2003
IR and IGF-1R double-knockout mice	DCM	Early-stage HF	Transcript (no change)	Laustsen *et al.,* 2007
		End-stage HF	Transcript (UP)	
Ventricular pacing in dogs	DCM-like	End-stage HF	Transcript (UP)	Zolk *et al.,* 2002
F vs NF human hearts	IDCM	End-stage HF	Protein (UP)	Nagueh *et al.,* 2004
F vs NF human hearts	IDCM	End-stage HF	Transcript (UP)	Zolk *et al.,* 2002
			Protein (UP)	

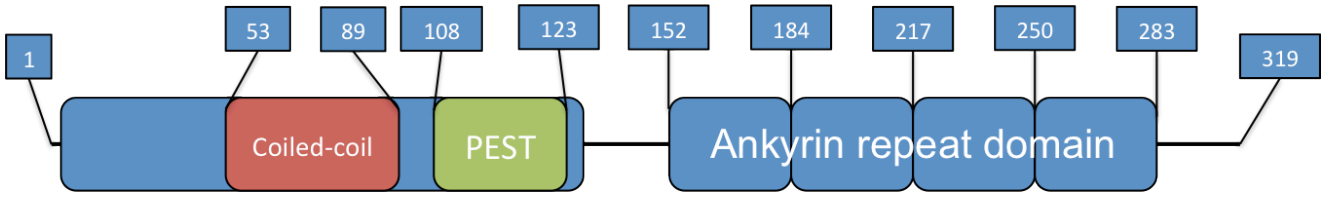
Abbreviations: MLP - muscle LIM-only protein; CASQ2 - cardiac calsequestrin; Gs-α -guanidin nucleotide-binding α; IR - insulin receptor; IGF-1R - insulin growth factor-1 receptor; VH - ventricular hypertrophy; DCM - dilated cardiomyopathy; IDCM - idiopatic DCM; F - failing heart; NF - non-failing heart.

1.1.2 Structural analysis of CARP

CARP, identified in 1995 (Chu *et al.,* 1995), consists of 319 amino acids (MM ~36kDa). *S*equence-based structure predictions of CARP indicate that the N-terminal fraction of this protein is largely unstructured (containing a PEST-like sequence and a short coiled-coil domain), while the C-terminal fraction folds into four highly conserved ankyrin repeats (AR) (Jeyaseelan *et al.,* 1997; Ishiguro N *et al.,* 2002) (*Fig.1.2A*). The short coiled-coil domain within the N-terminal fraction has been reported to induce homodimerization of CARP (Torrado *et al.,* 2004; Witt *et al.,* 2005), with split yellow fluorescent protein (BiYFP) and GST-pulldown assays indicating that the N-terminal region of CARP (residues 1-122) self-associates in an anti-parallel fashion (Lun *et al.,* 2014). The AR domain at the C-terminus of CARP is considered as the mediator of the CARP:titin N2A interaction (Lun *et al.,* 2014). CARP also has numerous potential modification sites for phosphorylation, glycosylation and myristoylation (Chu *et al.,* 1995; Jeyaseelan *et al.,* 1997).

A recent study of CARP structure (Ewan Craig, Honours Report, U.Liverpool, 2014) included additional prediction tools, namely, QUARK (Zhang *et al.,* 2012), I-TASSER (J Yang *et al.,* 2015) and MODELLER (Webb B and Sali A, 2014) to inspect the 3D-fold of the AR C-terminal region. AR motifs normally contain ~33 amino acids that fold into a helix-loop-helix α-hairpin structure, with each adjacent α-helix arranged in an antiparallel fashion (Breeden *et al.,* 1987; Tevelev *et al.,* 1996; Byeon *et al.,* 1998). The study hinted at the possible existence of 1-2 additional, likely distorted, AR repeats in the C-terminal fraction of CARP. The study also used the disorder predictor PONDR-fit (Li *et al.,* 1999) (http://www.disprot.org/pondr-fit.php) to better identify domain borders in CARP (*Fig.1.2B*). The result of those predictions served as guidance to design expression constructs for the current study.

A.



B.

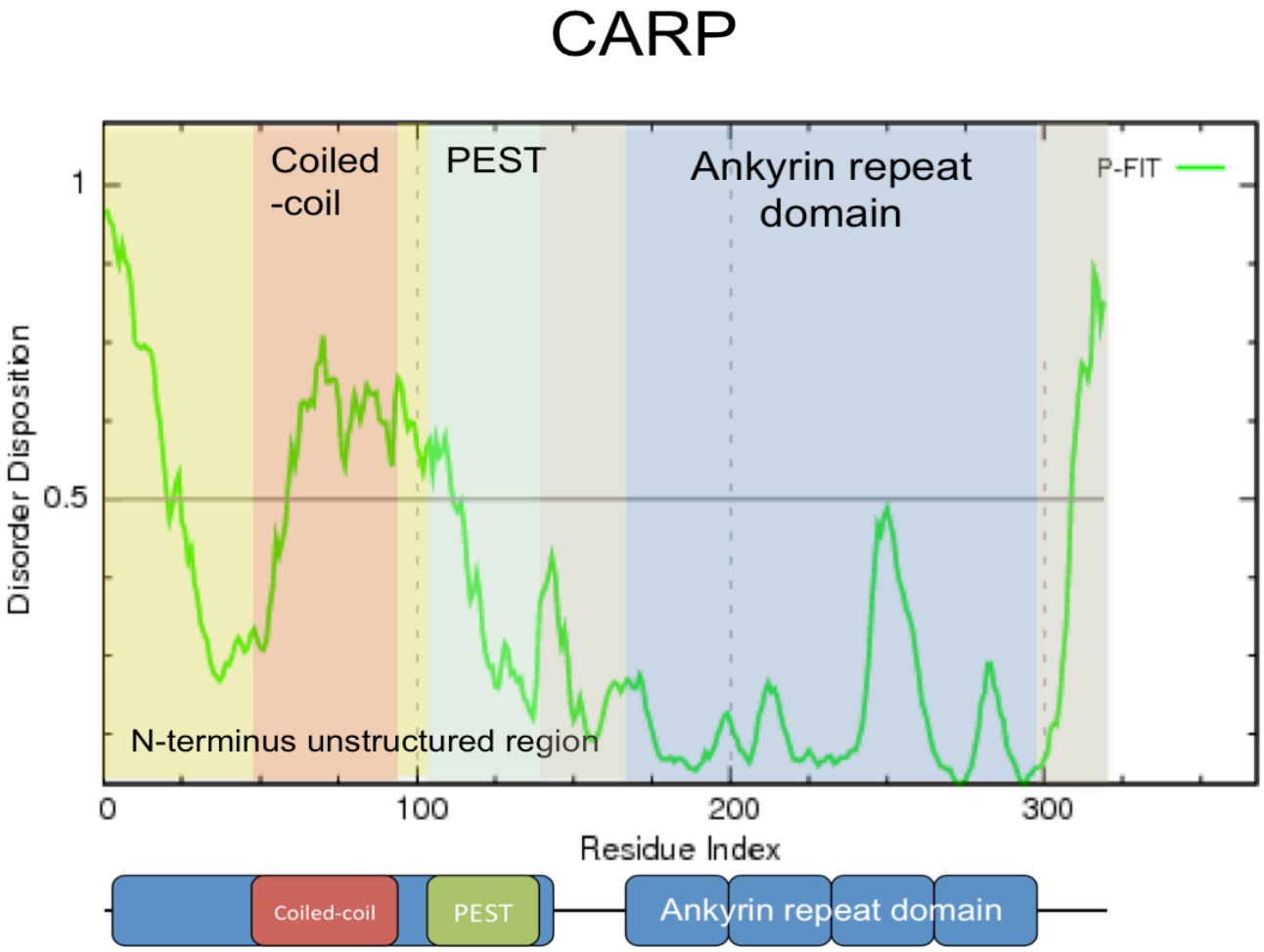


Fig1.2. **Structure prediction of CARP**

A) Predicted domain arrangement of CARP shows that the N-terminus region contains a coiled-coil domain and two PEST-like sequences and a C-terminal AR region (Jeyaseelan *et al.,* 1997; Mikhailov *et al.,* 2008).

B) CARP sequence analyzed using PONDR-fit disorder prediction, a score (y-axis) over 0.5 is considered disordered. The N-terminus of CARP (residue 1 to ~110) was predicted to be more than 50% disordered. However, the C-terminus AR domain gives a lower score, which means most of this region is likely to be structured, at exception of the C-terminal tail (Ewan Craig, Honours Report, U.Liverpool, 2014). (The prediction of a coiled-coil domain in a disordered fraction is not an uncommon result from current prediction tools; Anurag *et al.,* 2012).

1.2 Functions of the CARP:titin N2A interaction

1.2.1 Titin modulates the passive tension of the sarcomere

The sarcomere is the basic contractile unit of muscle. It is about 2-3μm long and has a diameter of 1-2μm. Sarcomeres are joined with each other end-to-end and build the myofibril (Squire JM, 1997). The sarcomere can be divided into several zones: Z-disk, I-band (isotropic in polarized light), A-band (anisotropic in polarized light) and M-line (*Fig.1.3*). The basic force generators of the sarcomere are the thin actin-based filament and the thick myosin-based filament (Hanson and Huxley, 1953; Squire JM, 1997). However, the sarcomere also contains several gigantic, passive filaments that are essential for the maintenance of its ultrastructure, passive mechanics and regulation. Among these, the myofilament titin is the third most abundant protein in muscle and the best-known member of the family of sarcoskeletal elastic proteins.

Titin is a large protein with a molecular weight that varies from 2970kDa to 3700kDa depending on isoform (Labei*t et al.,* 1995; Bang *et al.,* 2001). It spans a half sarcomere, from the Z-disk to the M-line (approx. 1.2 μm) (*Fig.1.3; Fig.1.4*). Titin is implicated in the generation of passive tension in the sarcomere, which is achieved by the stretch of its I-band region (Helmes *et al.,* 1999; LeWinter *et al.,* 2010).

Titin's I-band region, where CARP is thought to associate, contains different elements that form a molecular spring: two tandem Ig segments (proximal and distal), a PEVK-rich region and one or two unique elements (N2A and N2B) depending on isoform (Miller *et al.,* 2004; Taylor *et al.,* 2011) (*Fig.1.4*). Titin has two main isoforms in heart, called N2B and N2BA (Labeit *et al.,* 1995). The difference between these two isoforms is the structure of their I-band. The N2B isoform contains the N2B unique element and a short PEVK segment. However, the N2BA isoform contains the N2B element, the N2A region, a middle tandem Ig sequence and a long PEVK region (Miller *et al.,* 2004) (*Fig.1.4*). Another important difference between these two isoforms is the stiffness. Under stretch, the N2B isoform is stiffer than the N2BA. The regulation of the ratio of these two isoforms in heart is considered as an important pathway to the control of myocardium

stiffness (Cazorla *et al.,* 2000; Martin *et al.,* 2007). A reflection of the mechanical differences of these two isoforms and their significance for heart performance is that, their expression levels vary across species. In human, the N2BA:N2B ratio in the left ventricle is normally around 0.6, while the atria contain more N2BA titin. As a comparison, the mouse heart, which beats much faster than the human heart, contains over 95% of N2B isoform (Cazorla *et al.,* 2000; LeWinter *et al.,* 2010). Additional pathways to muscle stiffness regulation include the modification of titin through phosphorylation and glutathionylation (or even proteolysis) and the binding of sarcomeric partners, such as CARP the subject of this study.

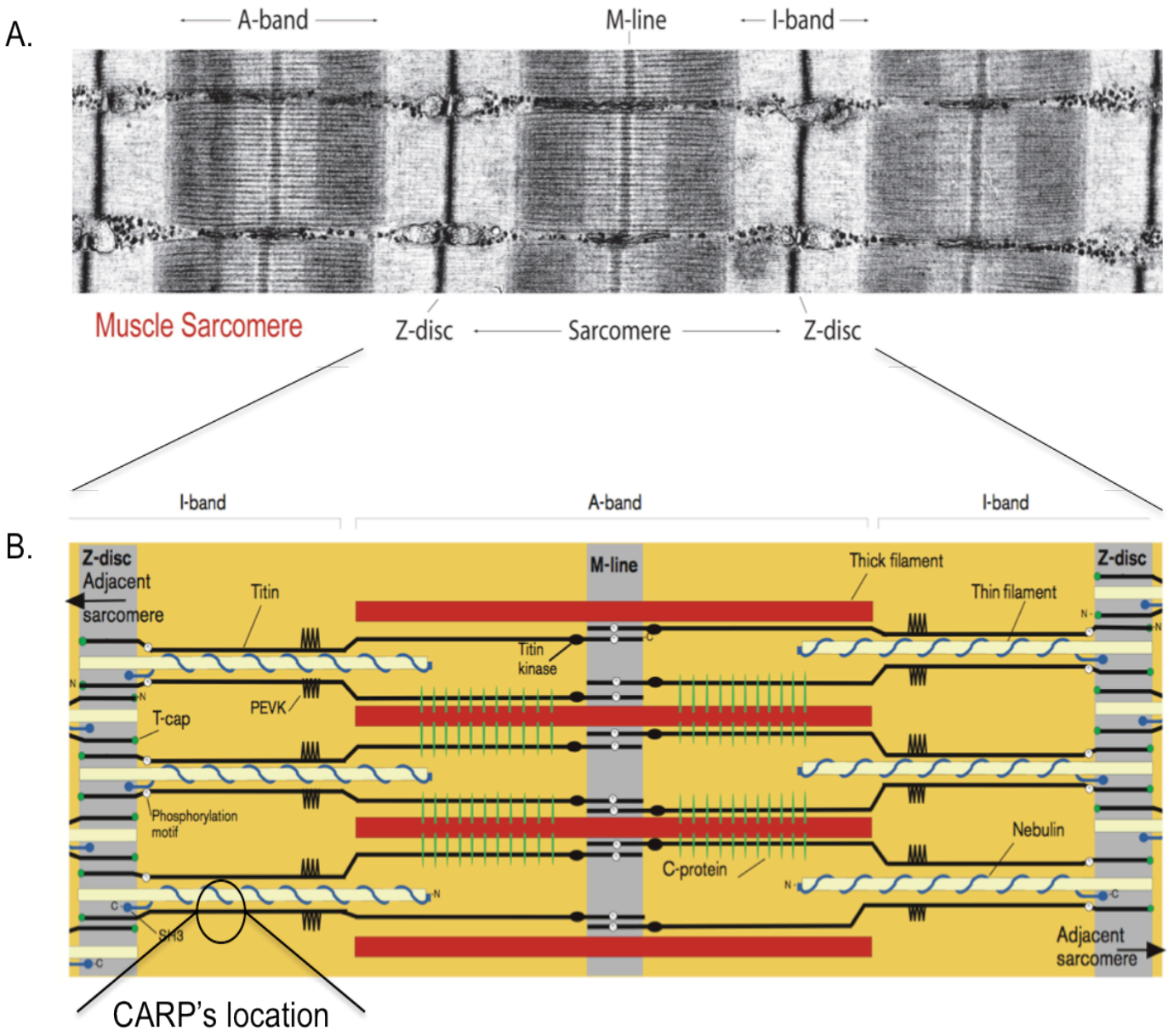


Fig.1.3 The muscle sarcomere

A) Electron micrograph showing the structure of the sarcomere. (Figure from Eric *et al.,* 2007).

B) The four primary filament systems of the sarcomere: actin (light yellow), myosin (red), titin (black), and nebulin (blue). The red band is the thick filament, which is mainly formed by myosin. The thin filament is formed by actin and nebulin. Titin spans the M-line and the Z-disk (half sarcomere). CARP is found at the I-band, bound to titin. (Figure from Gregorio *et al.,* 1999).

1.2.2 CARP binds to the N2A region of titin and modulates its stiffness

The N2A region of titin contains two unique and four Ig domains (*Fig.1.4*). Tests by yeast two-hybrid screening (Y2H) indicated that CARP interacts with titin N2A at the elastic unique domain (UN2A) between Ig80 and Ig81 (Miller *et al.,* 2003). Another measurement using BiYFP provided the same result and additionally found that CARP binds titin N2A via its C-terminal AR motif (105-319) (the N-terminus of CARP (residues 1-122) was unable to interact with the titin N2A region) (Lun *et al.,* 2014).

By binding to the cardiac-specific titin N2A region, CARP may serve as a muscle compliance modulator in the heart. Such stiffness modulation appears induced by inhibition of phosphorylation of the titin N2A region (Lun *et al.,* 2014), by PKA (or PKG), which increases titin's compliance and reduces the passive force of the sarcomere (Krüger and Linke, 2006; Krüger *et al.,* 2009; Hidalgo *et al.,* 2012). However, the exact phosphorylation site is not clear. Accordingly, myofibrils from knock-out mice for the three proteins of the MARP family (CARP, MARP and DARP) were found to have a high phosphorylation level and became more compliant as the length of the myofibril was increased, which supports the view that CARP can control the stiffness of titin by stopping phosphorylation of the N2A region (other members of MARP family can also interact with titin and may have the same function as CARP in skeletal muscle) (Barash *et al.,* 2007; Lun *et al.,* 2014).

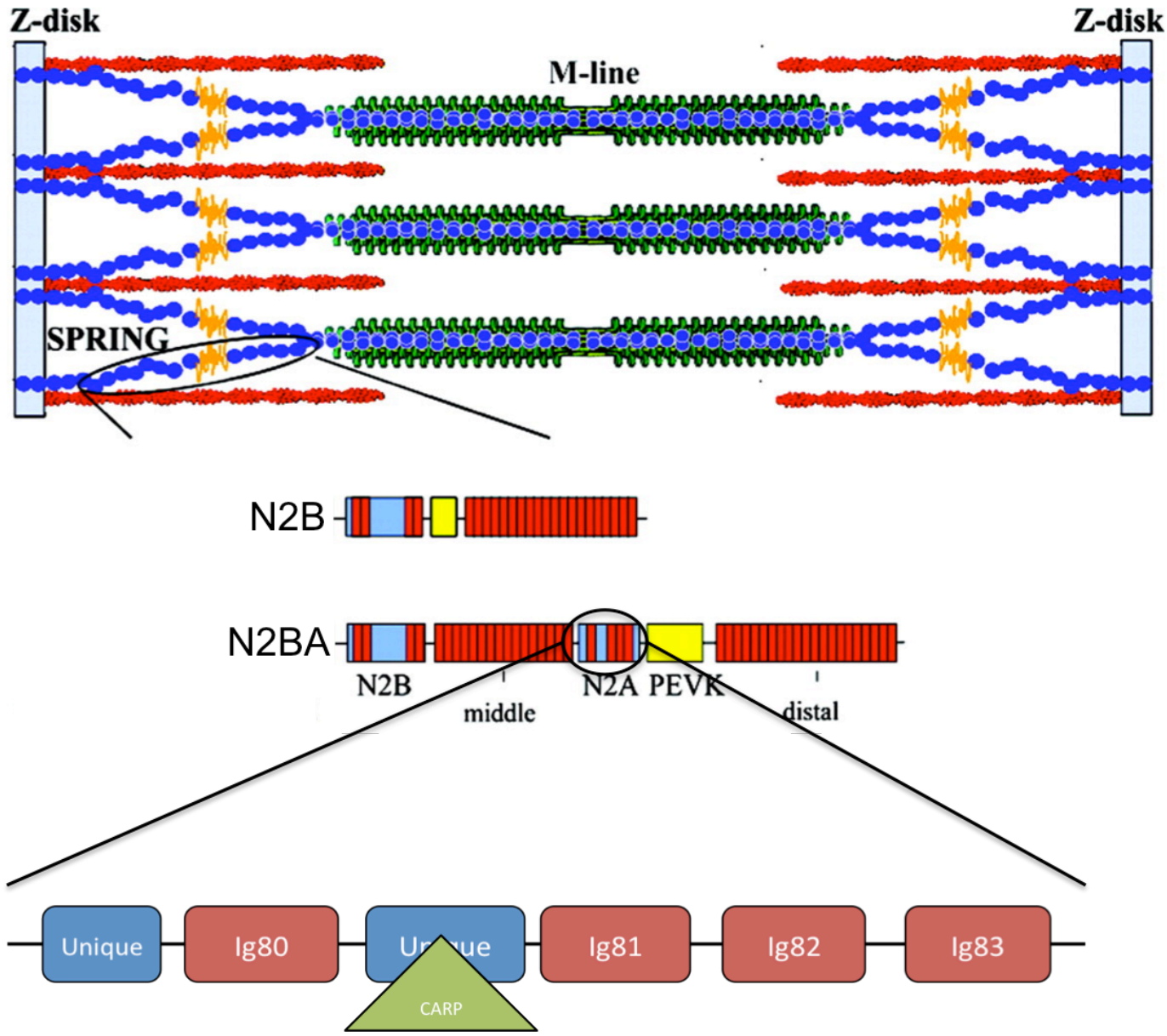


***Fig.1.4* Structure of two titin isoforms in the I-band**

Titin (blue beaded line) has two main isoforms in the heart and their I-band structure is different. The stiffer N2B isoform is shorter, with a short PEVK segment and a N2B unique element between two tandem Ig regions. The more compliant N2BA isoform is longer and the specific element contains both N2B and N2A region, an extended PEVK segment and an additional Ig repeats area. The N2A region contains four Ig (Ig80, Ig81, Ig82 and Ig83) and two unique domains, with the unique domain between the Ig80 and Ig81 considered as the binding site of CARP (Figure reproduced from Matthew *et al.,* 2011).

1.3 Aims

In this project, we aimed to characterize the CARP:titin-N2A interaction biophysically and set the ground for future structural studies. In this respect, we aimed to reveal the minimally interacting fraction of CARP and N2A that would yield high-affinity complexes suitable for study. The ultimate goal is to generate molecular knowledge that can guide future functional studies on this complex and its suitability for the development of therapeutic interventions to combat HF.

In the experimental plan, the first goal was to design viable constructs of CARP and titin N2A and establish their recombinant overexpression in bacteria as stable, soluble and pure samples. Subsequently, the samples would be used to generate and validate the formation of the complex *in vitro*. This would be followed by the biophysical characterization of samples, such as their oligomeric state, their secondary structure content and their stability using circular dichroism (CD), Nuclear Magnetic Resonance Spectroscopy (NMR), Differential scanning fluorimetry (DSF) and Size Exclusion Chromatography combined with Multi-Angle Laser Light Scattering (SEC-MALLS). Complexation between CARP and titin-N2A domain variants would be monitored by gel-filtration (GF), and the affinity of the binding (K_D) quantitated using Microscale thermophoresis (MST). If complexes could be isolated and purified, SEC-MALLS would be used to identify the stoichiometry of the binding. Crystallization trials of stable samples and their complexes would then follow, aiming to obtain protein crystals suitable for structural elucidation with a view to a future atomic characterization of the complex.

2. Methods

2.1 Expression clones

Expression clones for all proteins were provided by Dr. Julius Bogomolovas at Uniklinikum Mannheim (Germany), with the exception of the N2A-2 variant (described below).

Coding sequences (specified in Chapter three, *Fig.3.1; Fig.3.4*) from truncated variants of CARP (UnitProtKB Q15327) and domain variants of the titin N2A region (UnitProtKB Q8WZ42) were inserted into the pET-trx1a vector, with the exception of variants N2A-1 and N2A-2 that were inserted into the pET-trx1b vector. Both pET-trx1a and pET-trx1b vectors fuse a His_6-tag, a thioredoxin domain and a TEV (tobacco etch virus) protease cleavage site N-terminal to the inserted gene. However, the order of the tags is different, with the His_6-tag being in N-terminal position in vector pET-trx1a but in internal position in pET-trx1b (*Fig.2.1*).

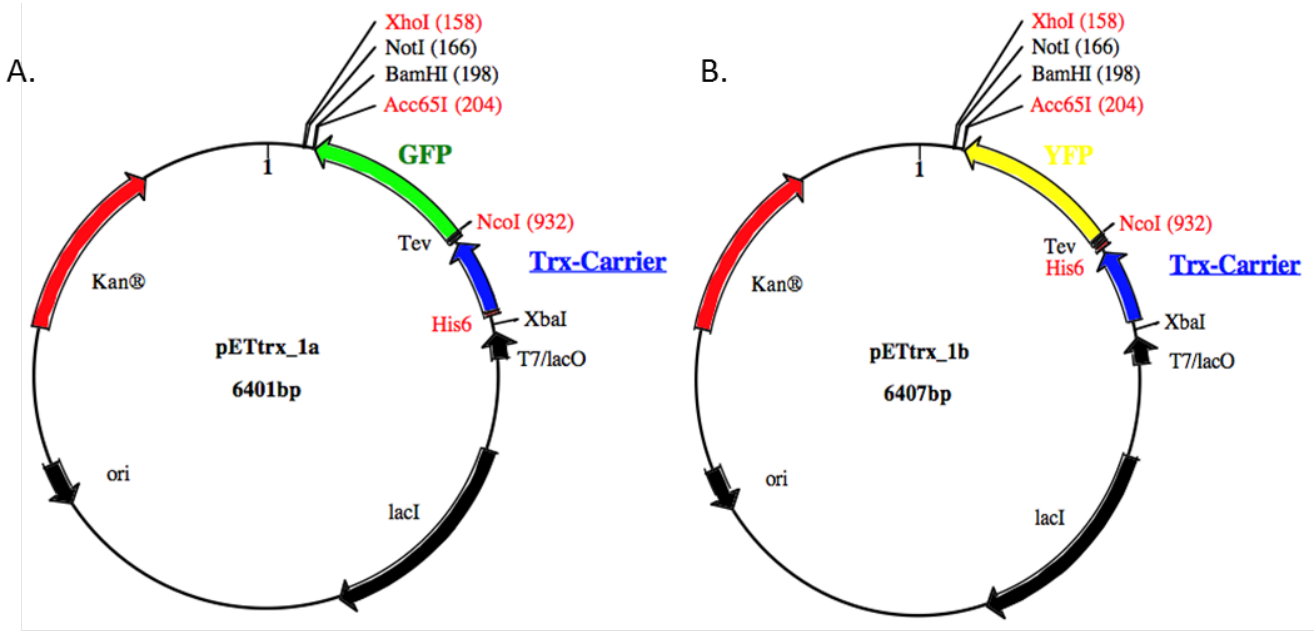


Figure reproduced from:
http://babel.ucmp.umu.se/cpep/web_content/pdf/vector%20maps/pETtrx1.pdf

***Fig.2.1* Expression vectors.**

A) The pET-trx1a vector, the GFP position (green) is replaced by the target gene.

B) The pET-trx1b vector, the YFP position (yellow) is replaced by the target gene.

The N2A-2 gene was made using PCR-amplifying techniques. The primers
Forward primer: 5'-CCA TGG GTC TGT TTT TTG TTA GCG AAC-3'
Reverse primer: 5'-GGT ACC TTA CTC AAT AAC GGT CAG TTT TGC-3'
were designed to add the restriction sites **NcoI** (CCATGG) and **KpnI** (GGTACC) to the gene. The enzyme used for PCR was KOD hot start enzyme (Merk Millipore), and the PCR reaction mixture contained:
1µl template (16.8ng/µl plasmid)
1µl enzyme, 2.5µl primers
5µl buffer, 3µl $MgSO_4$,
5µl ddNTP, 30µl water
The PCR reaction conditions are shown in *Table.2.1*.

Table.2.1 The reaction cycles of PCR.

Step	**Temperature**	**Time**	
Denature	95°C	2minutes	
Denature	95°C	20seconds	30 cycles
Annealing	59°C	10seconds	
Amplifying	70°C	15seconds	
Extending	70°C	10minutes	

Next, the gene was isolated using agarose gel electrophoresis and purified using a gel-extraction kit (BIOLINE). T4 ligase (New England Biolabs) was used to insert the gene into the vector using the ligase reaction mixture:
6µl vector (79.5ng/µl)
3.5µl PCR products (20.5ng/µl),
2µl 10x buffer, 1.5µl enzyme,
7µl dd water

The mixture was incubated at room temperature for 2 hours, and then used to transform *E.coli* strain DH-5α (Invitrogen) for plasmid amplification. Bacteria were grown at 37°C for 16-18h before harvest. The plasmid was extracted using a mini prep kit from MACHEREY-NAGEL and the gene concentration was estimated by measuring A_{260} using a nanodrop spectrophotometer (ND-1000, Labtech).

All plasmids were sequenced (Source Bioscience and GATC Biotech). Sequence data were translated using a translation tool (http://web.expasy.org/translate/) and compared with the wild type sequence (from Uniprot) by BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) before expression. Basic protein parameters, such as the molar mass (MM) of a monomer, the pI and the extinction coefficient were calculated with ProtParam (http://web.expasy.org/protparam/).

Bioinformatics tools, such as Psi-pred (http://bioinf.cs.ucl.ac.uk/psipred/) or Predict Protein (https://www.predictprotein.org) were used to predict secondary structure content from sequence data. PONDR-fit disorder prediction (http://www.disprot.org/pondr-fit.php) (Xue *et al.,* 2010) was used to predict low complexity areas in the proteins from their sequences.

2.2 Recombinant protein production

All proteins were expressed in *E. coli* strain Rosetta (DE3) (Merck Millipore). The bacteria was pre-cultivated in 10ml Luria Bertani (LB) media (Merck Millipore) supplemented with 25µg/ml kanamycin and 34µg/ml chloramphenicol for 16-18 hours at 37°C, 180rpm. The pre-cultures were used to inoculate 1L cultures that were further grown in LB media (with 25µg/ml kanamycin and 34µg/ml chloramphenicol) at 37°C, 160rpm, until an OD_{600}=0.6. At this point, protein expression was induced with 0.5mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Bioline) and cultures further grown overnight at 18-22°C, 160rpm. Bacteria were harvested by centrifugation at 4225*g* for 20-30 minutes and stored at -80°C until further use. Upon thawing, the cells were lysed with a pressure cell homogenizer (Standsted) at 394MPa in affinity chromatography (AC) buffer (composition described in *Table.2.2*) containing an EDTA-free protease inhibitor cocktail (Roche) and 1mg DNAse (Sigma-Aldrich). The lysate was clarified by centrifugation at 38724*g* for 30 minutes. The supernatant was removed and filtrated with a 0.22µM syringe filter (Merck Millipore).

The purification of all proteins (except GFP-$CARP^{106-319}$) from supernatants followed three consecutive steps: AC, reverse affinity chromatography (RAC) and size exclusion chromatography (SEC). The GFP-$CARP^{106-319}$ sample was purified

applying only AC and gel-filtration (GF). The column used for AC and RAC was a Histrap™ HP 5mL (GE healthcare). For GF, several columns were employed depending on sample and whether preparative or analytical amounts were required: HiLoad Superdex 200 16/60, HiLoad Superdex 75 16/60, HiLoad Superdex 75 26/60 and HiLoad Superdex 75 10/300 GL (all from GE healthcare). The buffers used in each purification step are summarized in *Table2.2*. All buffers were filtered with 0.22μM filter paper before use and, in addition, GF buffers were degassed. All purification steps were performed in a ÅKTA Explorer chromatography system (GE healthcare). Chromatographs were first analyzed using Unicorn Manager (GE healthcare) and then exported to Excel (Microsoft).

Table.2.2 Buffers used throughout this work

Name of Buffer	**Conditions**
Affinity chromatography (AC)	25mM HEPES pH=7.5, 300mM NaCl
AC elution	25mM HEPES pH=7.5, 300mM NaCl, 0.5M imidazole
Gel-filtration (GF)	25mM HEPES pH=7.5, 100mM NaCl
Phosphate buffer (for CD)	20mM phosphate pH=7.5

Finally, protein purity was monitored by SDS-PAGE, with 6% stacking gel and different resolving gels (12% to 18%) depending on the MM of the protein. After purification, protein samples were concentrated by ultrafiltration (Merck Millipore) and the concentration estimated by measuring A_{280} with a nanodrop spectrophotometer (ND-1000, Labtech). Samples were stored at 4°C until further use.

To roughly estimate the MM of protein samples, GF calibration curves were made using aprotinin (6.5kDa), ribonuclease A (13.7kDa), carbonic anhydrase (29kDa), ovalbumin (44kDa), conalbumin (75kDa), aldolase (158kDa) and ferritin (440kDa).

2.3 Circular dichroism (CD) spectroscopy

The secondary structure content of $CARP^{106-319}$ and UN2A variants was estimated using CD performed on an 1100 CD spectrometer (JASCO). Proteins that had been stored in GF buffer were diluted 100 times with phosphate buffer (*Table.2.2*) to a

final concentration of 0.18mg/ml for CARP and 0.29mg/ml for UN2A. The spectral λ range of the measurements was 185nm-260nm using a 1mm cell (HELLMA 110-QS). All CD data were acquired three times and their average calculated and used for further analysis.

The secondary structure content of both $CARP^{106-319}$ and UN2A was calculated from CD data using Dichroweb (http://dichroweb.cryst.bbk.ac.uk). Five analysis algorithms were tested (CONTINLL, SELCON3, CDSSTR, VARSLC and K2D) against four databases that fitted the spectral range (details in section 3.3.2).

For thermal denaturation, sample conditions, cells and spectral ranges were as described above. $CARP^{106-319}$ was assayed in the temperature range 25°C to 80°C, with data collected two times at each integral degree and the average calculated. The temperature range for UN2A was 20°C to 95°C and the data acquired as before. The temperature change rate for both measurements was 1°C/min with an incubation time of 90 seconds per temperature step. Thermal denaturation curves were calculated at λ=208nm (λ=220nm for UN2A) and the melting temperature (Tm) obtained from the first derivative.

2.4 Size Exclusion Chromatography combined with Multi-Angle Laser Light Scattering (SEC-MALLS)

The MM of $CARP^{106-319}$, N2A-1, N2A-2, UN2A GFP-$CARP^{106-319}$ and various of their complexes were determined by SEC-MALLS on an ÅKTA pure (GE Healthcare) connected to a premier Multi-Angle static Light Scattering (MALLS) detector and a differential Refractive Index (dRI) detector (DAWN HELEOS-II and Optilab T-rEX; Wyatt Technology). The proteins were in GF buffer that had been three times filtrated (0.22μM filter) and degased before use. The column used was HiLoad Superdex 200 10/300 (GE Healthcare) for GFP-$CARP^{106-319}$ and the complex between GFP-$CARP^{106-319}$ and N2A-2, and HiLoad Superdex 75 10/300 (GE Healthcare) for all other tests. Column and MALLS machine were washed in GF buffer at a 0.2ml/min flow rate overnight and then the flow rate changed to 0.75ml/min for about 1 hour. The protein sample was injected at a volume of 200μl-500μl. The total amount of protein was approximately 1mg. The system

was calibrated using bovine serum albumin (BSA), immediately prior to the measurement of target proteins. All data were first processed and analyzed using ASTRA software (Wyatt Technology) and later exported into Excel (Microsoft).

2.5 Nuclear Magnetic Resonance Spectroscopy (NMR)

Purified N2A-1 and UN2A samples were studied using NMR by Dr. Igor Barsukov. The concentration of N2A-1 was 3mg/ml. The sample had been dialyzed overnight in phosphate buffer at pH=6.5 supplemented with 50mM NaCl, 3mM β-ME and studied by H^1-NMR at 600MHz. UN2A was concentrated to 15mg/ml in GF buffer and the sample examined by H^1-H^1-NMR (NOESY, Nuclear Overhauser Effect Spectroscopy) at 600MHz. The temperature of both tests was 25°C.

2.6 Microscale thermophoresis (MST)

The MST technique was applied to quantitate the binding (K_D) of $CARP^{106-319}$ (tagged with GFP) and several N2A variants (Wienken *et al.,* 2010). Samples were in GF buffer at initial concentrations of 80μM for GFP-$CARP^{106-319}$; 1.06mM GFP (negative control); 3.85mM Ig81; 0.54mM UN2A; 0.27mM N2A-1; and 0.61mM UN2A-Ig81. The experiment was performed at the University of Sheffield, using a Monolith NT.115 instrument (NanoTemper-Technologies), and the wavelength of the infrared (IR) laser is 1475nm +/- 15nm. The data were collected two times, with 20% MST power and 40% MST power, and at 10% LED power. Four types of capillaries were tested (standard, hydrophilic, hydrophobic and premium types).

2.7 Differential scanning fluorimetry (DSF)

DSF data were recorded on a real-time PCR system (Applied Biosystem) in the temperature range 25°C to 99°C. Protein samples were in GF buffer and at a concentration of 50μM. The concentration of the dye (Sypro-orange dye) is 50 times (50x). Each sample was measured in triplicate. The obtained data were first processed and analyzed using Stepone™ software V2.1 (applied biosystem) and later exported into Excel (Microsoft).

2.8 Crystallization

The crystallogenesis of $CARP^{106-319}$, UN2A, Ig81, and the complex made by $CARP^{106-319}$ and UN2A-Ig81 was trialed using a nanovolume dispenser crystallization robot (Innovadyne) on 96-well Intelliplates (Art Robbins Instruments) and using a vapour-diffusion, sitting drop setting at 19°C. For the screening of chemical media, the following commercial crystallization screens were used: JCSG-plus, Structure Screen 1+2, PACT premier and MORPHEUS from Molecular Dimensions; PEGRx, NATRIX and SALTRX from Hampton Research; WIZARD and CRYO from Emerald Bioststems.

The optimization of initial crystallites of $CARP^{106-319}$ obtained from commercial preparations used 6 custom-made, optimization screens (*Fig.3.33*) assayed at room temperature and/or 4°C.

2.9 Collection of X-ray diffraction data for domain I81

Data were collected on beamline i04-1 of the Diamond synchrotron (Didcot, Oxford) at a wavelength of 0.9282 Å, using a PILATUS 6M-F detector. Data were recorded for a total rotation of 360° in frames of non-overlapping 0.25° rotation width. Crystals were prepared for cryo-crystallography by vitrification at 100k in cryo-protection medium (20% ethylene glycol; 20% isoproanol; 0.2M $MgCl_2$; 0.1M Tris pH 8.5). Subsequently, they were mounted on a 0.1mm litho-loop (Molecular Dimension), stored in a puck pre-cooled in liquid nitrogen and shipped to the synchrotron in the frozen state. Data (from an optimal rotation wedge showing low radiation damage) were processed using the XDS/XSCALE software suite (Kabsch W, 2010).

3. Results

3.1 Overproduction of CARP as a soluble protein in E.coli

3.1.1 Construct design of CARP$^{106-319}$

Expression in *E.coli* and purification of full-length CARP (work by Dr. Julius Bogomolovas, Uniklinikum Mannheim) showed that the resulting sample is not stable and undergoes rapid degradation into multiple species as revealed by SDS-PAGE. It was hypothesized that the degradation might particularly derive from the N-terminus of the protein, which is predicted to be disordered (*Fig.1.2*). The N-terminus of CARP is not thought to be required for the interaction with titin N2A (*Lun et al., 2014*), therefore it was removed (residues 1-105) to create an N-terminally truncated CARP construct, CARP$^{106-319}$, viable for recombinant overexpression in the current study (*Fig.3.1*).

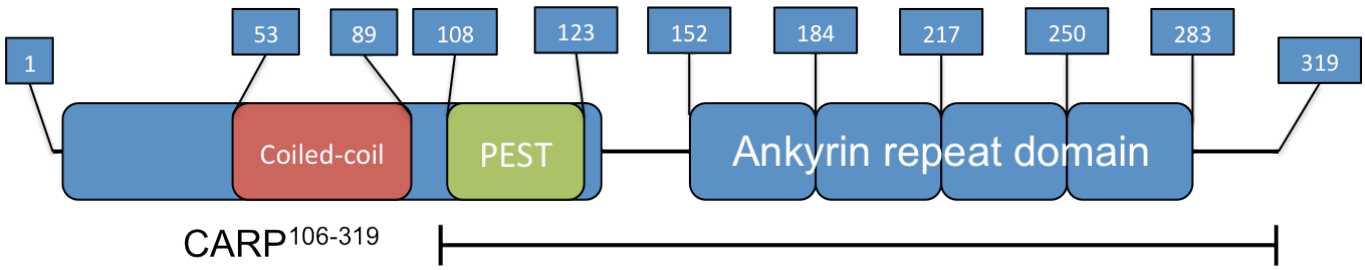
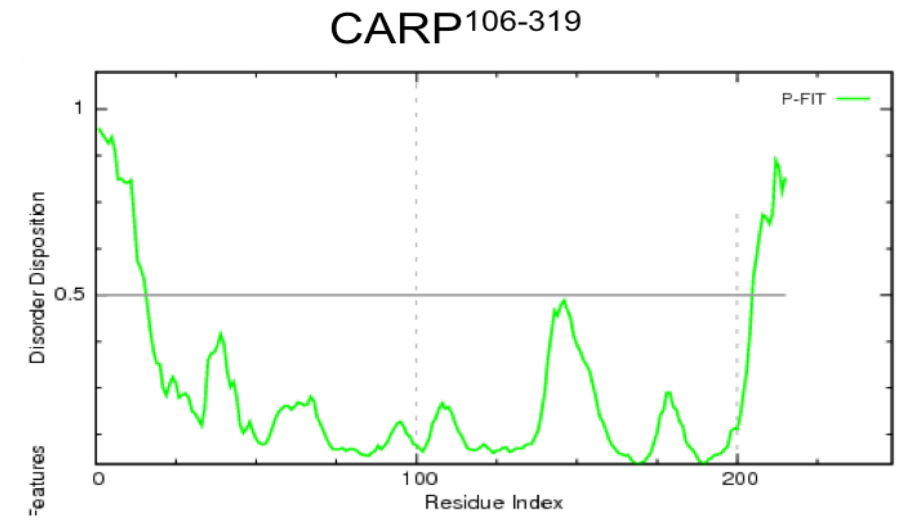


***Fig.3.1* CARP$^{106-319}$ constructs**

This CARP construct spans the PEST and AR domain (CARP; UnitProtKB Q15327).

The CARP$^{106-319}$ variant retains short stretches of allegedly disordered tail sequences that flank the AR region N- and C-terminally (approx. 16 and 11 residue long, respectively). The disordered character of these regions was predicted from sequence data using PONDR-fit (Xue *et al.,* 2010) and Psi-pred (http://bioinf.cs.ucl.ac.uk/psipred/), which report on low-complexity regions and secondary structure content, correspondingly (*Fig 3.2*). This analysis suggests that apart from the tails, the rest of the C-terminal CARP domain is well-structured, containing approximately 50% helical structure. This is in agreement with the expected AR protein fold. (A construct that lacked the hypothesized unstructured tails was generated by Dr. J. Bogomolovas and found to be notably less stable. Thus, it was excluded from the current study).

A.



B.

Fig.3.2 **The prediction result of CARP$^{106\text{-}319}$ secondary structure content**

A) The CARP$^{106\text{-}319}$ sequence analyzed using PONDR-fit disorder prediction. A score (*y*-axis) over 0.5 indicates predicted disorder. The result suggests that ~85% of this protein is ordered.

B) Prediction of secondary structure content using Psi-pred. The result predicts that the structure is largely helical (103/215 aas in helical conformation, amounting to 48% of the sequence).

3.1.2 Overproduction and purification of CARP$^{106-319}$

Overexpression of CARP$^{106-319}$ was established in *E. coli* and the resulting protein purified by AC and SEC. Typically, the expression yield of soluble CARP$^{106-319}$ in *E.coli* was 7mg of pure sample per 1L culture (purity was monitored using SDS-PAGE, *Fig.3.3*). The volume of exclusion (Ve) of this sample in SEC referenced to marker proteins (see Methods) suggested that CARP$^{106-319}$ has an approximate MM of 50kDa, corresponding to a dimer (the MM of the polypeptidic chain calculated from its sequence is 24.1 kDa). This result was not anticipated, as the CARP$^{106-319}$ construct was expected to be a monomer after removal of the N-terminal coiled-coil domain. This result shows the AR region can dimerize in the absence of the N-terminal coiled-coil domain.

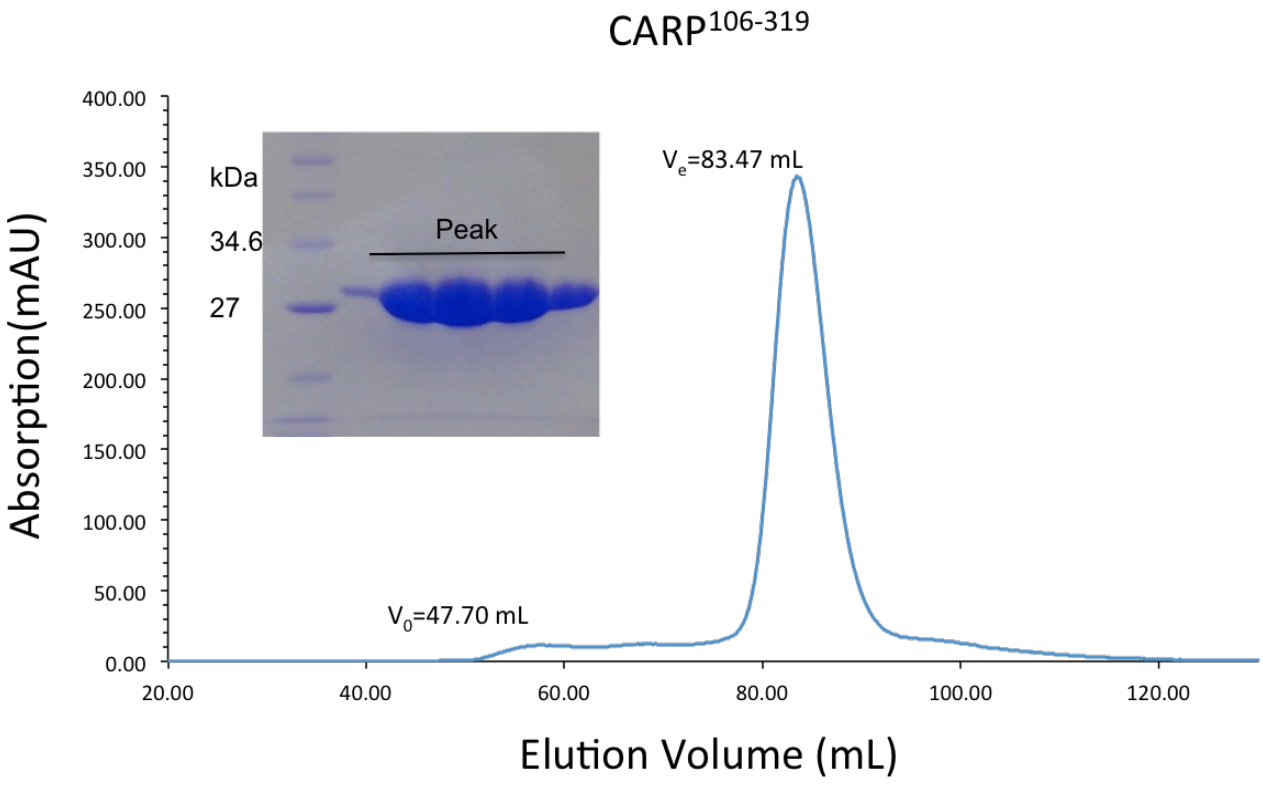


***Fig.3.3* Size exclusion chromatogram and SDS-PAGE of CARP$^{106-319}$**

Chromatogram (A_{280}) obtained on a Superdex 200 16/60 column. The lanes in the SDS-PAGE correspond to the elution fractions in the central part of the peak (which approximates a symmetric gaussian).

3.2 Recombinant overproduction of domain variants of titin N2A in E.coli

3.2.1 Construct design of titin N2A

The N2A element is a differentially spliced segment of titin that consists of two unique sequences and four Ig domains (*Fig.1.4*). The unique sequence between domains Ig80 and Ig81 (UN2A; 110 aa long) is thought to contain the CARP binding site (Miller *et al.* 2003). However, UN2A was believed to be intrinsically poorly ordered judging from sequence-based predictions (*Fig.3.5*). Thus, in this study, constructs containing the flanking Ig80 and Ig81 domains were designed, as it was speculated that these robust and highly-soluble domains might aid the production and storage of UN2A. As it was not possible to reliably predict the beginning of domain Ig80, a variant including the preceding unique sequence was also constructed. In addition to assist production, these construct variants would permit evaluating the contribution of the several N2A domains to CARP binding. Thus, seven constructs spanning various domain components of the titin N2A region were made and used in this project (*Fig.3.4*).

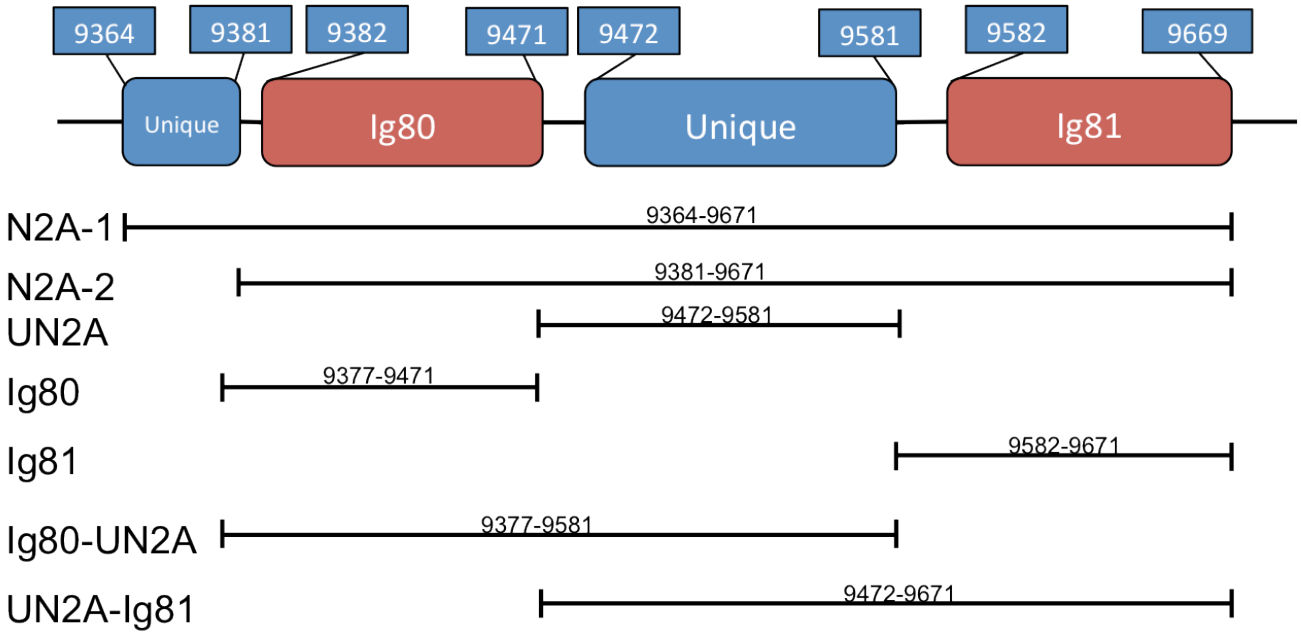


***Fig.3.4* Titin N2A constructs**

A total of seven constructs of the N2A region were made with diverse domain boundaries. The numbers above indicate the amino acid number (titin UniprotK entry Q8WZ42).

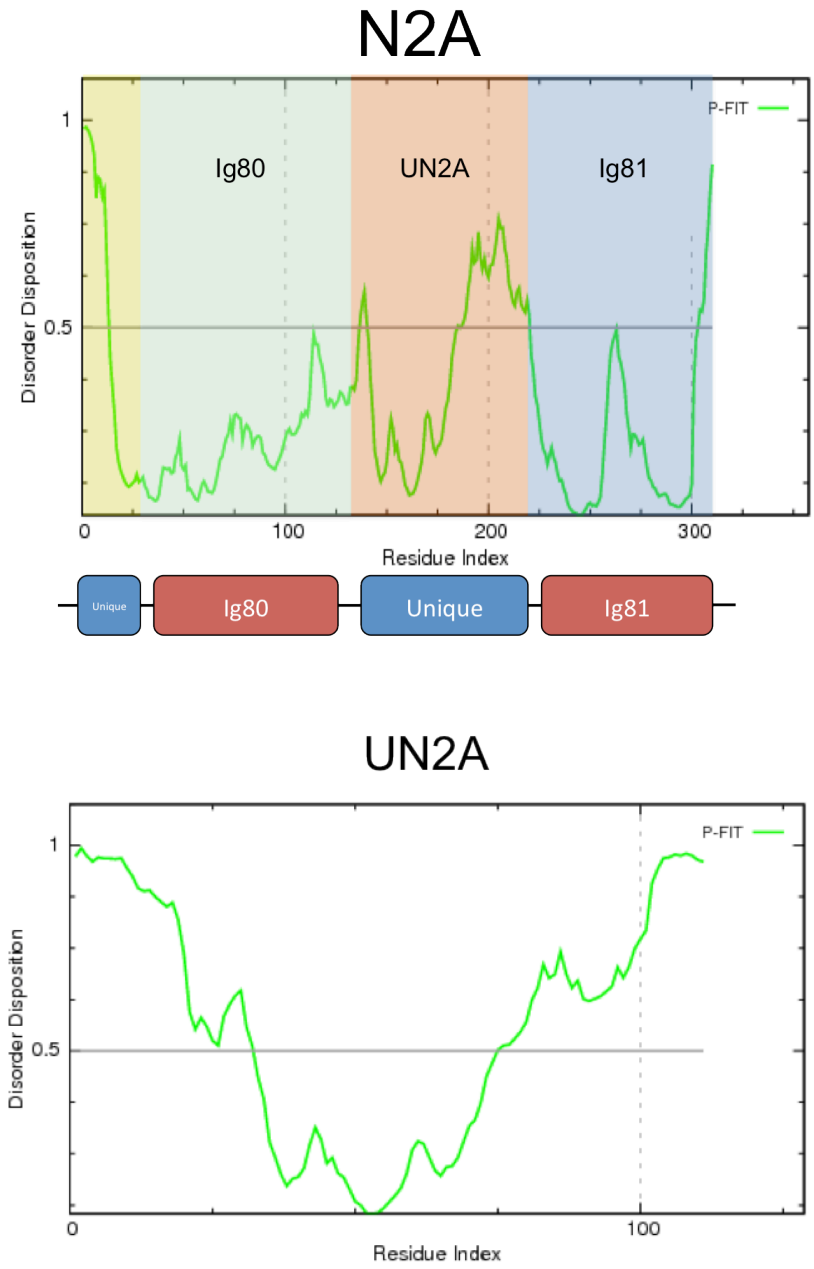


***Fig.3.5* Disorder prediction of titin N2A sequences**

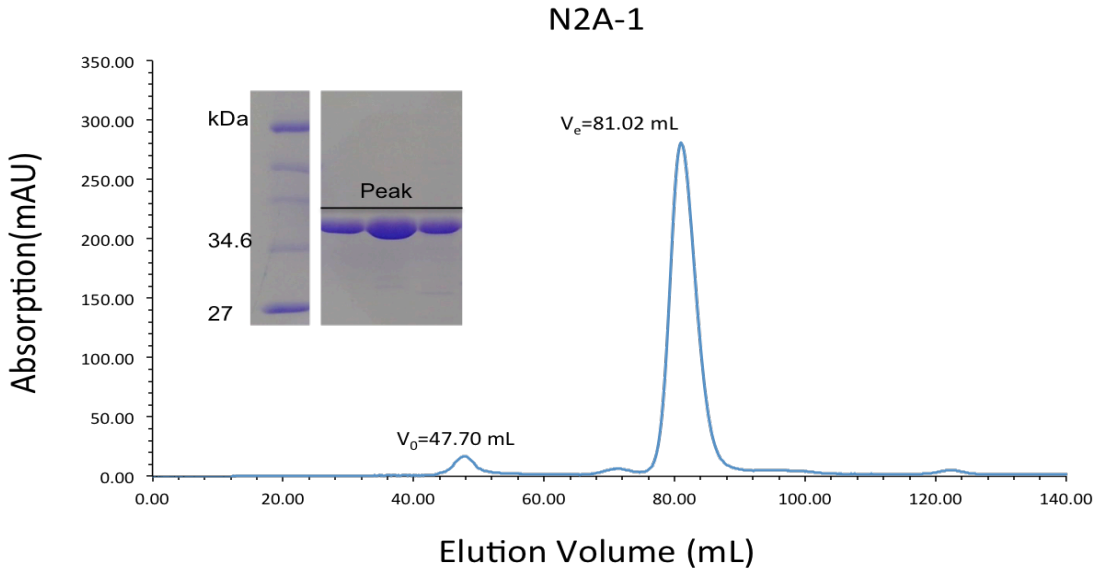
N2A-1 and UN2A sequences analyzed using PONDR-fit disorder prediction. A score (y-axis) over 0.5 indicates disorder. The result suggests that ~50% of UN2A is disordered. Ig80 and Ig81 are ordered.

3.2.2 Overproduction and purification of titin N2A

All constructs of titin N2A variants were expressed recombinantly in *E.coli*. Typically, 8mg of N2A-1, N2A-2 and Ig81, 10mg of UN2A-Ig81 and over 30mg of UN2A were obtained from 1L culture. Ig80 and Ig80-UN2A proteins were insoluble and were no longer pursued in this study.

The purification steps of all titin N2A proteins were as for CARP$^{106-319}$. The size exclusion chromatogram shows that N2A-1 and N2A-2 elute at Ve indicative of MM of 62kDa and 41kDa, respectively (*Fig.3.6*). The oligomeric state of these two proteins could not be inferred reliably from these values as the MM of their polypeptidic chains calculated from sequence data is 35.6kDa for N2A-1 and 33.9kDa for N2A-2).

A.



B.

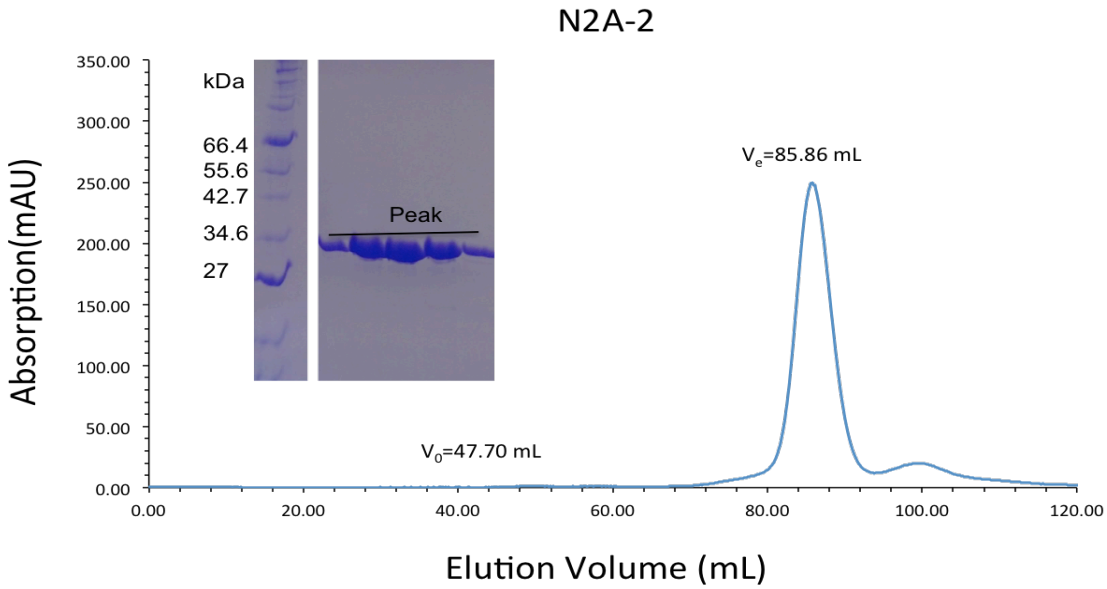


***Fig.3.6* Size exclusion chromatogram and SDS-PAGE of N2A-1 and N2A-2**

Chromatograms obtained on a Superdex200 16/60 column. The curve shows the A_{280}. The lanes in the SDS-PAGE correspond to the elution fractions in the central part of the peak (which approximates a symmetric gaussian). **A)** Result of N2A-1. **B)** Result of N2A-2.

The UN2A sample elutes in GF at a Ve indicative of approx. MM of 26kDa (*Fig.3.7*), corresponding to a dimer (the MM of the polypeptidic chain calculated from its sequence is 13.1 kDa). Interestingly, the SDS-PAGE of UN2A revealed an upper band, present in low quality. Further investigation using mass spectrometry (performed by Dr. Mark Willkinson) showed that this is the protein with a high degree of methionine oxidation.

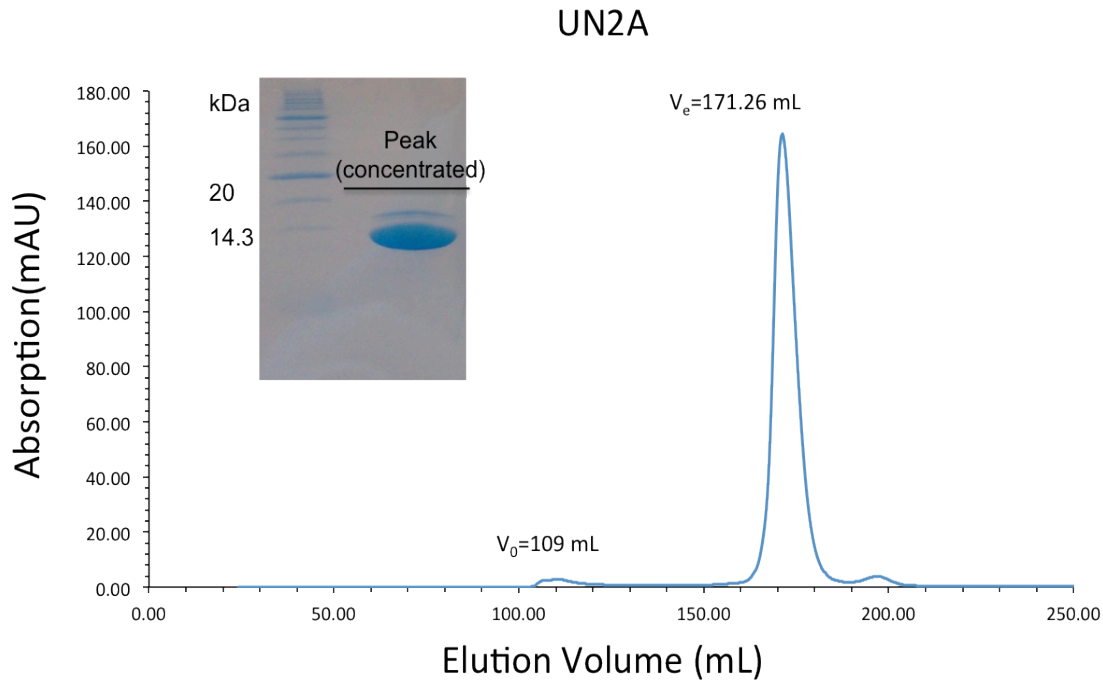
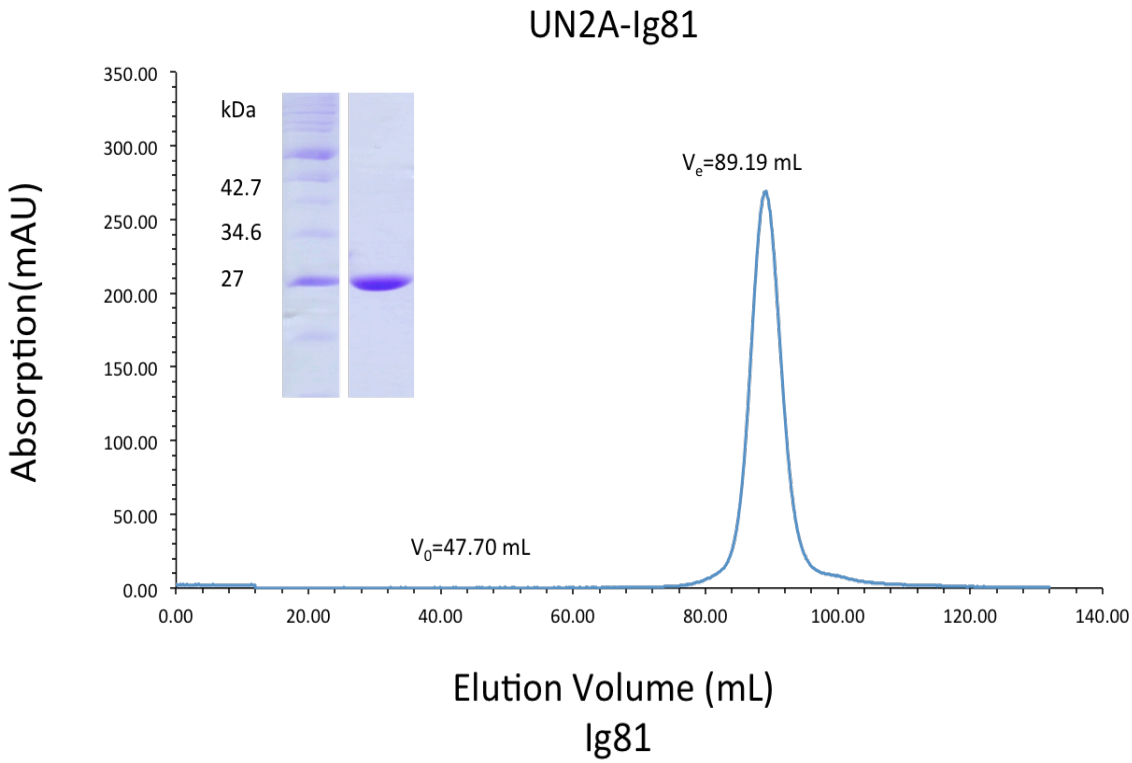


***Fig.3.7* Size exclusion chromatogram and SDS-PAGE of UN2A**

Chromatogram (A_{280}) obtained on a Superdex 75 26/60 column. The lanes in the SDS-PAGE correspond to the elution fractions in the central part of the peak (which approximates a symmetric gaussian). The upper band in the SDS-PAGE is UN2A with a high degree of methionine oxidation.

Samples UN2A-Ig81 and Ig81 elute at Ve indicative of approx. MM of 31kDa and 21kDa (*Fig.3.8*), corresponding to dimeric species of Ig81 but the oligomeric state of UN2A-Ig81 could not be inferred (the MM of the polypeptidic chain calculated from its sequence is 23.6kDa for UN2A-Ig81 and 10.5kDa for Ig81).

A.



B.

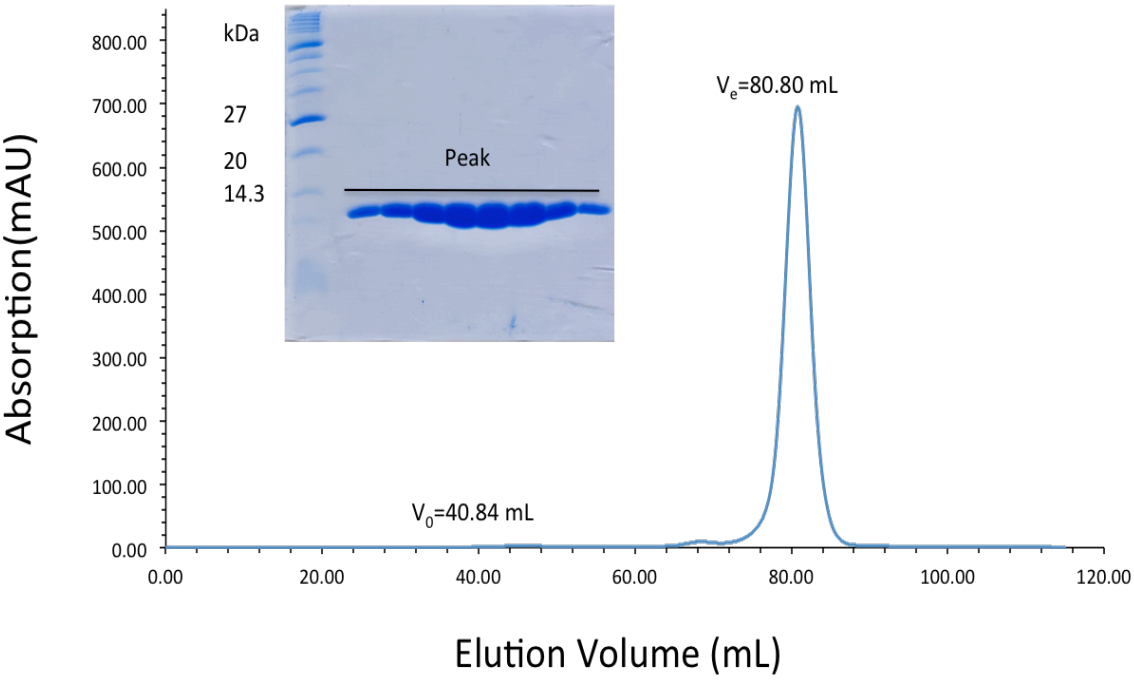


Fig.3.8 **Size exclusion chromatogram and SDS-PAGE of UN2A-Ig81 and Ig81**

The curve shows the A_{280}. The lanes in the SDS-PAGE correspond to the elution fractions in the central part of the peak (which approximates a symmetric gaussian). **A)** Chromatogram of UN2A-Ig81 obtained on a Superdex200 16/60 column. **B)** Chromatogram of Ig81 obtained on a Superdex75 16/60 column.

3.3 Biophysical characterization of isolated CARP$^{106-319}$ and titin N2A variants

3.3.1 SEC-MALLS measurement of CARP$^{106-319}$ and titin N2A

The MM calculated from size exclusion chromatograms is not accurate or necessarily reliable as elution volumes are commonly affected by hydrodynamic factors (e.g. the shape of the protein) and potential interactions with the resin. Therefore, the MM of CARP$^{106-319}$ and three of the titin N2A samples were measured using SEC-MALLS.

Analysis of CARP$^{106-319}$ by SEC-MALLS yielded a measured MM of 48.6kDa. As the theoretical MM of this protein is 24.1kDa, this result indicates that CARP$^{106-319}$ is a dimer (*Fig.3.9*). This agrees with previous conclusions from GF data, confirming that CARP dimerization is not solely due to the N-terminal coiled-coil domain but that the AR domain may also play a role in self-interaction.

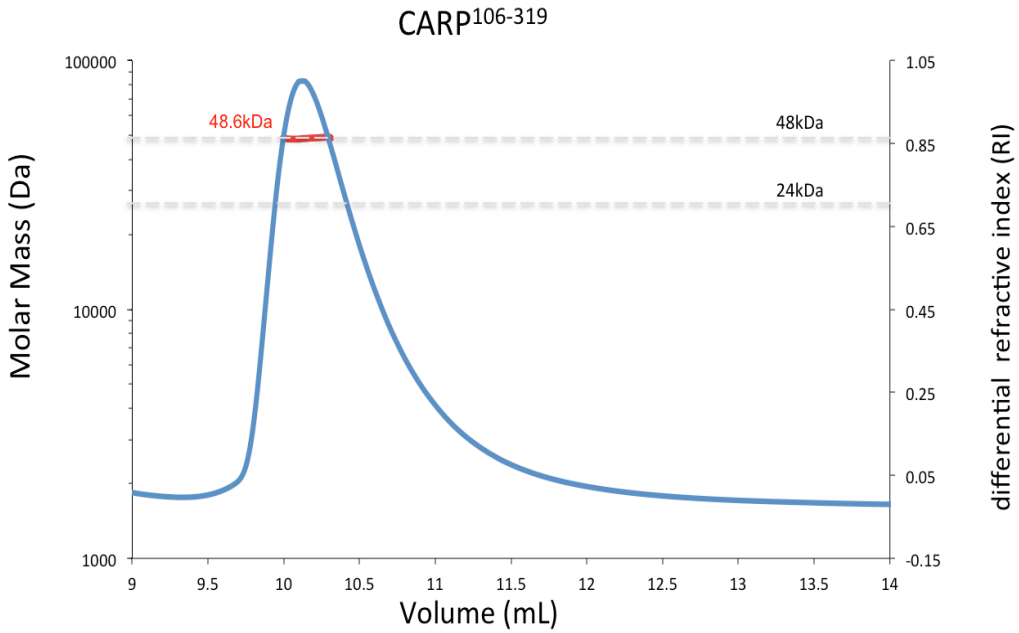
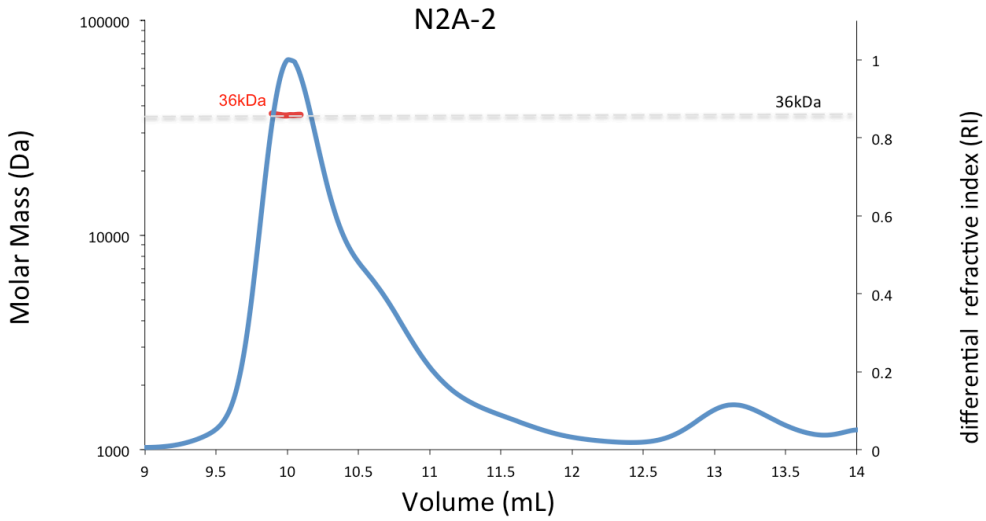


***Fig.3.9* SEC-MALLS analysis of CARP$^{106-319}$**

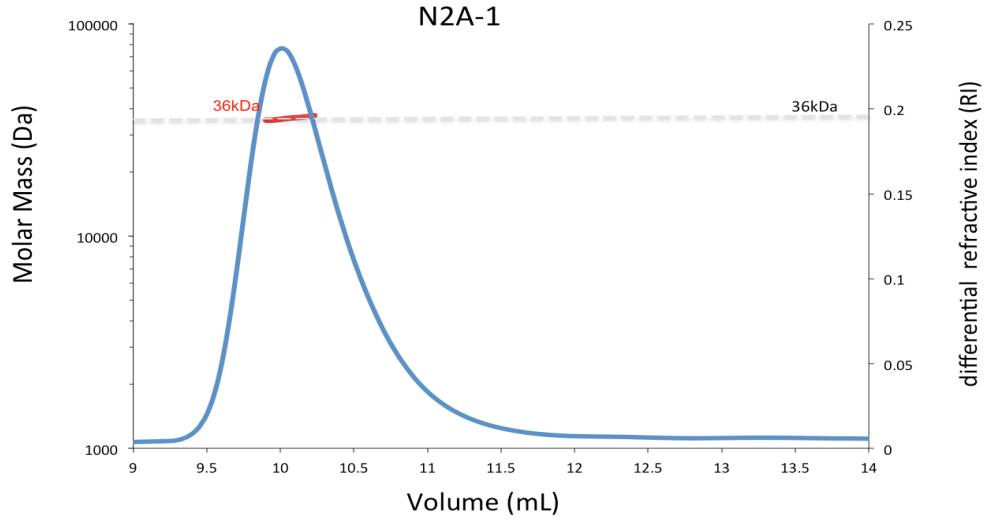
Normalized refractive index (blue) and measured MM (red) are plotted against elution volume using a Superdex 75 10/300 size exclusion column. The average MM measured was 48.6kDa. Theoretical MM for monomer and dimers of this sample are indicated using a dashed grey line.

Analysis of both N2A-1 and N2A-2 using SEC-MALLS revealed a MM of 36.0kDa. As the theoretical MM of these two proteins is 35.6kDa for N2A-1 and 33.9kDa for N2A-2, the results indicate that both of these constructs are monomeric. Additional analysis of UN2A yielded a MM value of 13.3kDa. Considering that the theoretical MM of UN2A is 13.1kDa, this protein is also a monomer. This supports the hypothesis that all the titin N2A constructs are monomeric (*Fig.3.10*).

A.



B.



C.

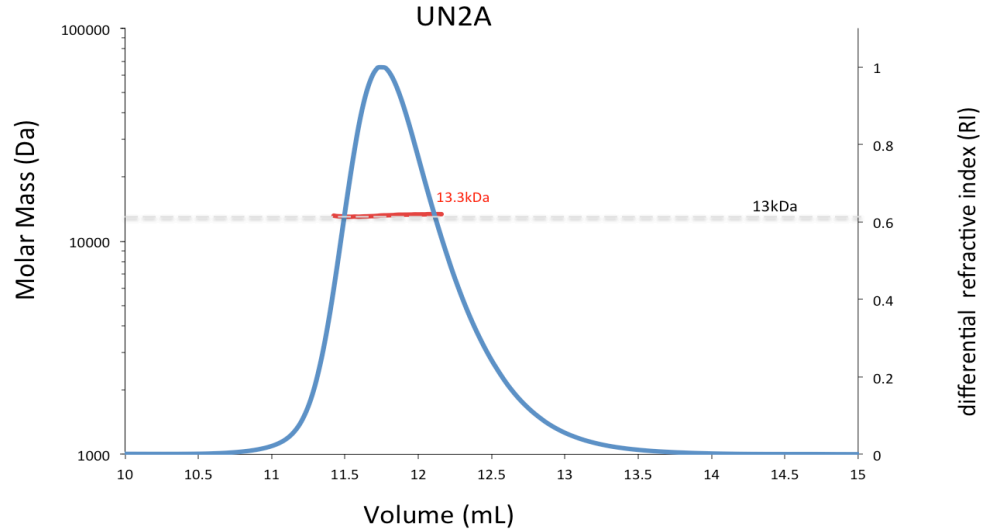


***Fig.3.10* SEC-MALLS analysis of titin N2A proteins**

Normalized refractive index (blue) and the average MM per volume unit (red) are plotted against elution volumes using a Superdex 75 10/300 size exclusion column. Theoretical MM for monomeric forms of the diverse samples are indicated using a dashed grey line.

All of the MMs calculated from sequences and measured by SEC-MALLS are shown in *Table.3.1*.

Table.3.1 The summary of MMs calculated by sequence and measured by SEC-MALLS

Constructs	Theoretical MM	SEC-MALLS MM	Oligomeric state
CARP$^{106-319}$	24.1kDa	48.6kDa	Dimer
N2A-1	35.6kDa	36.0kDa	Monomer
N2A-2	33.9kDa	36.0kDa	Monomer
UN2A	13.1kDa	13.3kDa	Monomer

3.3.2 CARP$^{106-319}$ and titin N2A samples have well structured and thermally stable folds

The measurement of secondary structure content is a reliable way to establish whether a protein is folded and, thus, suitable for atomic structural characterization. In this study, the CD technique was employed for this purpose, with secondary structure content calculated from CD data using the server Dichroweb (http://dichroweb.cryst.bbk.ac.uk) (Whitmore and Wallace, 2004; Whitmore and Wallace, 2008). This software generates reconstructed CD data that aim to reproduce the experimental CD spectrum by using a database of secondary structures and an analysis program that ultimately yields percentage estimates of the α-helix, β-sheet, turns and unordered component fractions of the protein. The accuracy of the calculation is given by the similarity of the experimental and reconstructed CD spectra according to a calculated Normalized Root Mean Squared Deviation (NRMSD), where a NRMSD larger than 0.05 means the reconstructed data do not fit the experimental data acceptably.

The Dichroweb server provides five analysis programs: CDSSTR (Johnson, 1999), CONTINLL (Sreerema *et al.,* 1999), SELCON3 (Van *et al.,* 1990), VARSLC (Compton *et al.,* 1986) and K2D (Andrade *et al.,* 1993). Among them, CDSSTR was chosen to calculate secondary structure content in this study because this algorithm yielded the best fit between calculated and experimental spectra (NRMSD=0.031 for CARP$^{106-319}$, and NRMSD=0.01 for UN2A). CONTINLL and SELCON3 algorithms gave NRMSD values above 0.12. VARSLC considers the spectral range starting at

178nm, which is larger than the range of the experimental data in this study (from 185nm to 260nm) and, thus, could not be used here. The spectral range of K2D is 205-240nm, which is insufficient to cover the experimental range, and this algorithm also gave a NRMSD>0.15. Thus, CDSSTR was deemed here as the most suitable mode of analysis.

In addition to the different algorithms, Dichroweb provides seven different secondary structure databases. Their applicability depends on the wavelength range of the available experimental data and referenced proteins, with four of the databases matching the spectral range in this study (Set 3, Set 4, Set 6 and Set 7) (Wallace and Janes, 2009). All four databases for CDSSTR were tried and the results were basically identical.

The CD-based calculation of secondary structure content for CARP$^{106\text{-}319}$ indicated that this protein contains 41% helical and 13% β-turn components (*Fig.3.11*). This result is in good agreement with the Psi-pred prediction that this protein contains ~48% of helical structure (*Fig.3.2*). Interestingly, Psi-pred predicts 12 α-helices, instead of the 8 that would be strictly needed to form 4 AR motifs. This agrees with a previous estimation of 1 (less probable 2) additional AR repeats in CARP (Ewan Craig, Honours Report, U.Liverpool, 2014). If only 8 α-helices were present in this fraction of CARP, the measured α-helical content would have approximated 30-32% (judging from secondary structure predictions *Fig.3.2*). Thus, the CD data here obtained bring now experimental support to the presence of additional helical elements, specifically they appear to confirm the presence of 5 AR total in CARP (as this corresponds to roughly 40% helical content).

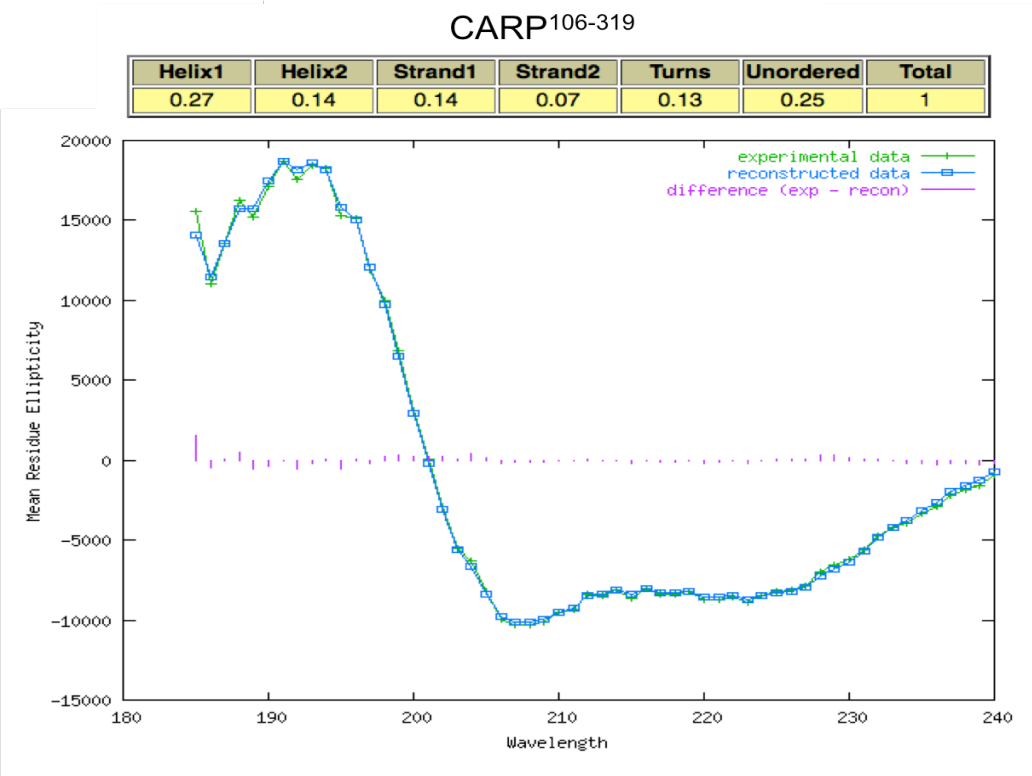


Fig.3.11 **The CD analysis result of CARP$^{106\text{-}319}$**

Helix1 and Helix2 are both α-helix, strand1 and strand2 are both β-sheet. The green curve is the CD experimental data, reconstructed data is shown in blue and the purple line is the difference between experimental data and reconstructed data. The NRMSD of this measurement is 0.031.

The CD calculation of UN2A indicates that this is a highly helical protein (62% helical content) (*Fig.3.12*). This result was unexpected as the initial working hypothesis (due to a strongly held belief in this scientific community) was that UN2A was largely unstructured. To further explore this result, we performed a secondary structure prediction (Psi-pred). The prediction was in good agreement with CD data (*Fig.3.13*) suggesting ~70% of the sequence in helical conformation. Further supporting these results, we also found similarity between a central UN2A sequence and the folded region of the Unique Cardiac Myosin Binding Protein-C Motif (71% sequence conservation), that forms a three-helix bundle and whose structure has been determined by NMR (Howarth et al., 2012). Interestingly, in the unique cardiac MyBPC Motif is the helical region adjacent to a poorly ordered N-terminal segment, potentially being reminiscent of the situation in UN2A (*Fig.3.14*). At a later stage, we validated this unexpected observation using 2D-NMR (see section 3.3.3).

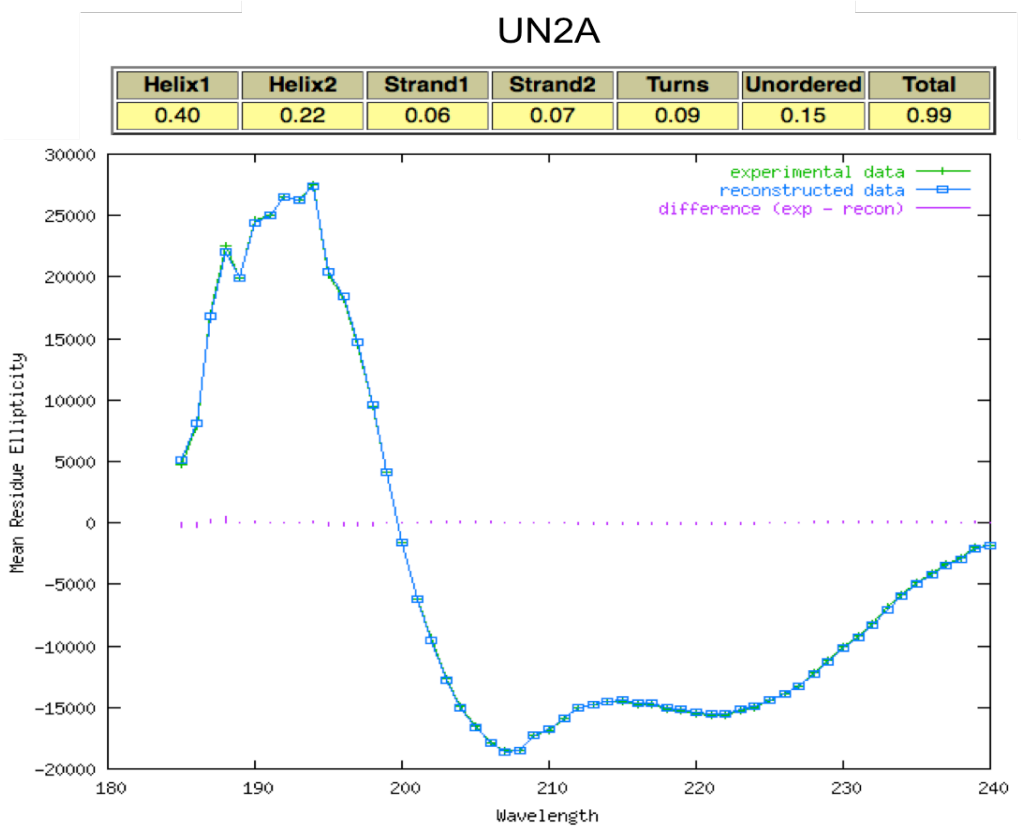


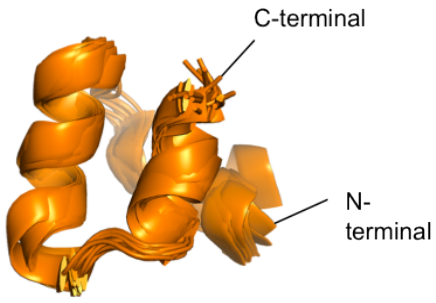
Fig.3.12 **CD analysis of UN2A**

Helix1 and Helix2 are both α-helix, strand1 and strand2 are both β-sheet. The green curve is the CD experimental data, reconstructed data is shown in blue and the purple line is the difference between experimental data and reconstructed data. The NRMSD of this measurement is 0.01.

Fig.3.13 **Secondary structure prediction for UN2A**

Prediction of UN2A's secondary structure content using Psi-pred. The result predicts that the structure is largely helical (80/110 residues are predicted in helical conformation amounting to 72% of the total sequence).

A.



B.

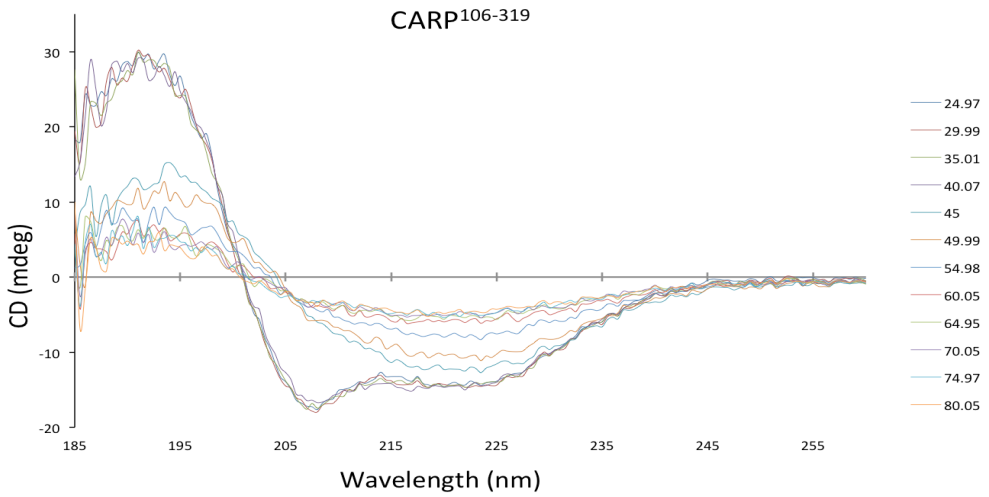
```
UN2A:    EIDIMELLKNVDPKEYEKYARMYGITDFRGLLQAFELLKQSQ
         E D+ E+L+    P EYE+ A  +G+TD RG+L+  + +KQ +
MyBPC3:  EEDVWEILRQAPPSEYERIAFQHGVTDLRGMLKRLKGMKQDE
```

Fig.3.14 **A)** NMR structure of the Unique Cardiac Myosin Binding Protein-C Motif (PDB entry 2LHU); **B)** sequence alignment of UN2A and the MyBPC motif, where conservation is indicated and the helical segments are highlighted in yellow.

After asserting that CARP$^{106-319}$ and UN2A are well-structured samples and that appear stable during storage (monitored by SDS-PAGE, *Fig.3.31*), next the robustness of the fold was assayed by thermal denaturation. For this, the changes in the CD spectra upon temperature increase were used to calculate the melting temperatures (Tm) of CARP$^{106-319}$ and UN2A.

For CARP$^{106-319}$, as the most obvious change in the spectra is observed at 208nm, the change in CD signal at this wavelength was used to calculate a Tm=43°C (*Fig.3.15*).

A.



B.

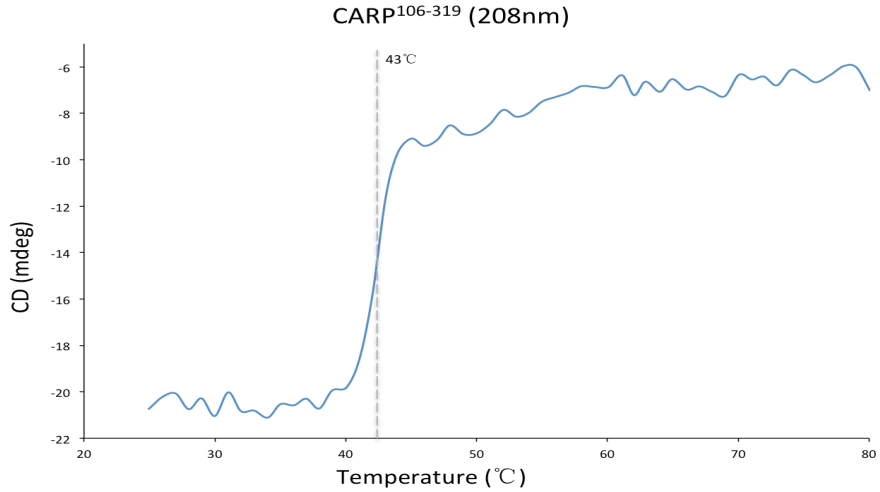


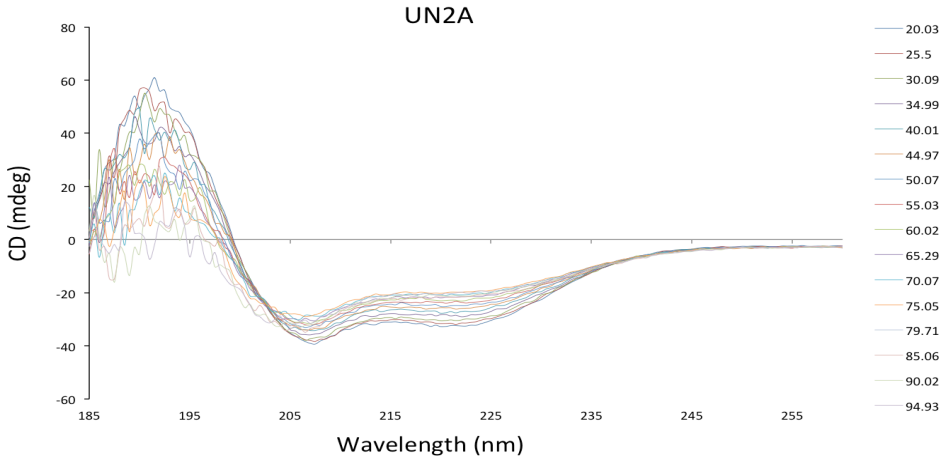
***Fig.3.15* CD-monitored thermal denaturation of CARP$^{106-319}$**

A) The CD spectra of CARP$^{106-319}$ at different temperatures (for visual simplicity, spectra at 5°C intervals are shown). The corresponding temperatures are shown on the right.

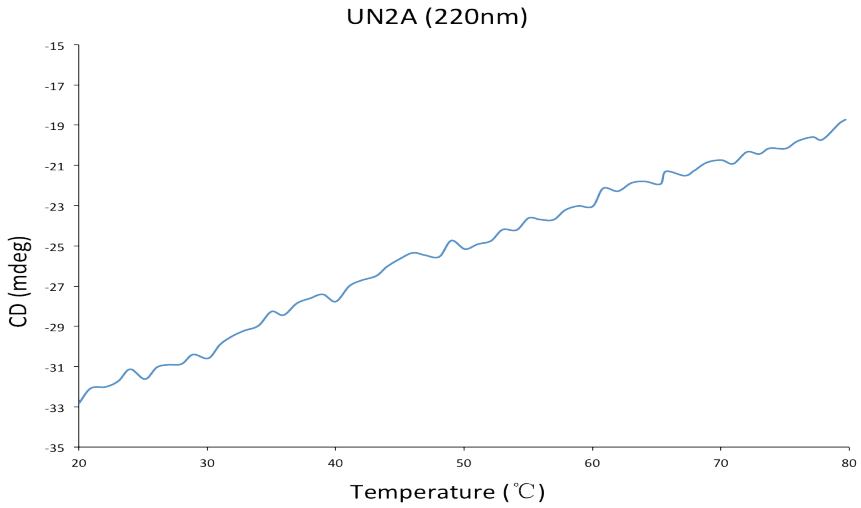
B) Temperature denaturation curve of CARP$^{106-319}$ was recorded at a wavelength of λ=208nm. This protein shows a Tm=43°C.

The CD-monitored thermal denaturation of UN2A indicated that this protein is highly stable, retaining its secondary structure after incubating at 95°C for 90 seconds (*Fig.3.16*). An additional CD measurement of UN2A was made, after the protein was boiled at 99°C for ~10 minutes. Also here, UN2A retained its secondary structure, indicating that its fold is unusually robust, being well beyond the levels commonly observed in human intracellular proteins and approximating the stability of proteins characteristic of bacterial extremophile organisms (*Fig.3.16*)

A.



B.



C.

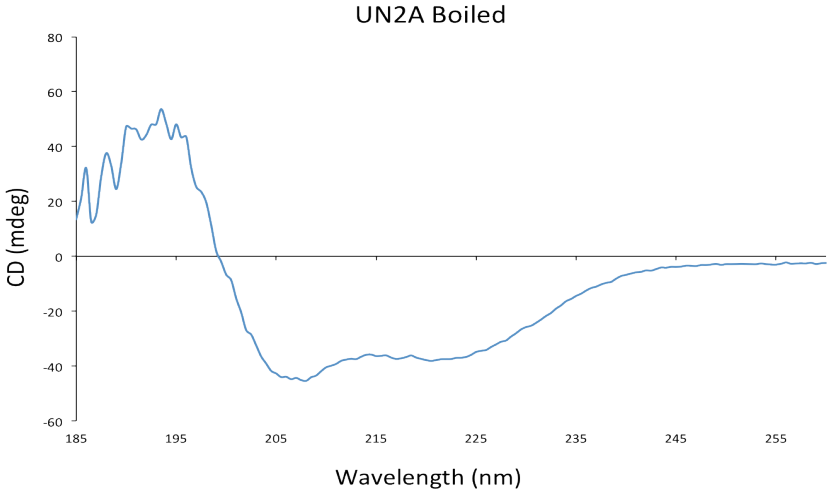


Fig.3.16 **CD-monitored thermal denaturation of UN2A**

A) The CD curve of UN2A at different temperatures (spectra at 5°C intervals are shown). Corresponding temperatures are shown on the right.

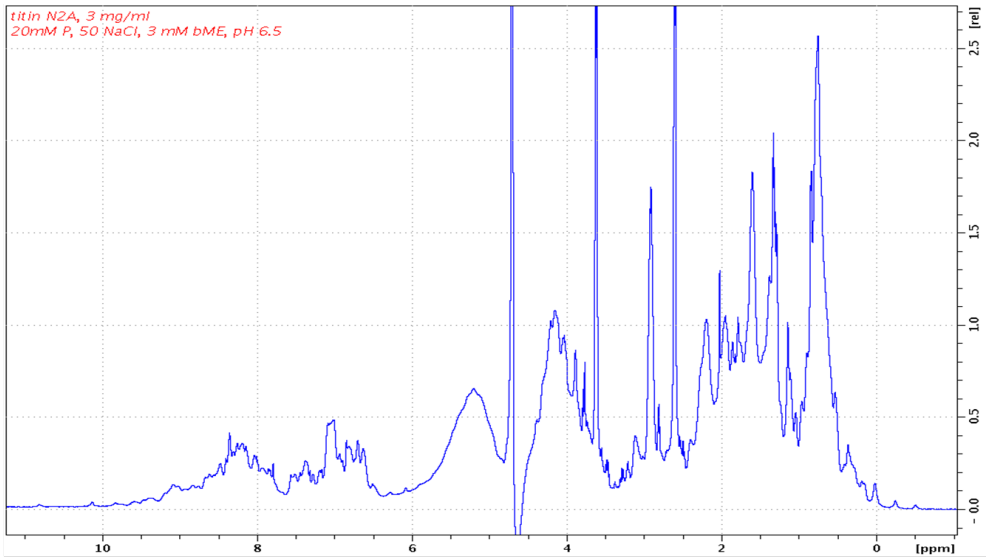
B) Temperature denaturation curve of UN2A recorded at a wavelength of λ=220nm. The Tm of this protein could not be calculated from these data.

C) The CD curve indicates that UN2A retains its structure after boiling.

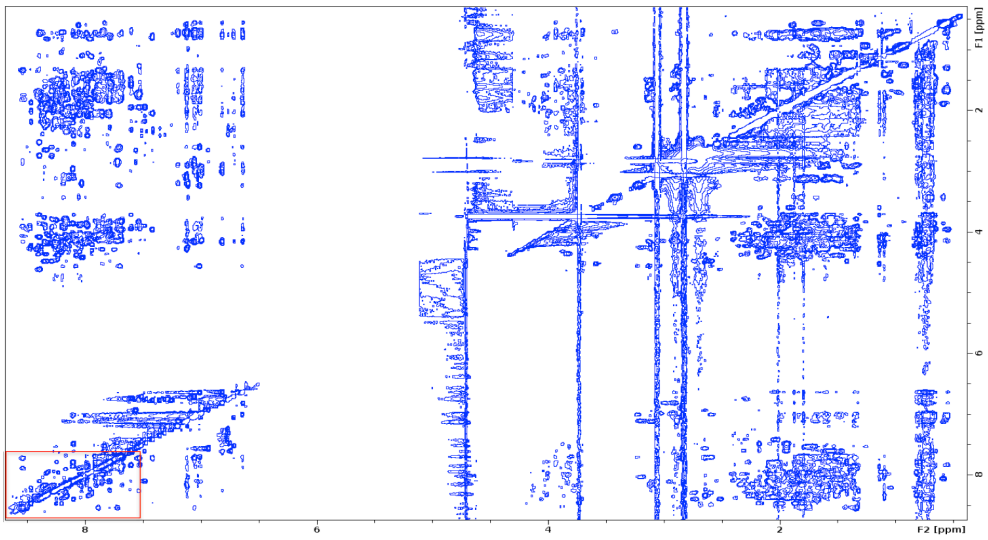
3.3.3 NMR spectroscopy of CARP$^{106-319}$ and titin N2A

NMR spectroscopy is a highly sensitive method that reports on protein folding with higher accuracy than CD. Hence, UN2A, without isotope labelling, was analyzed using H^1-H^1-NMR (NOESY, Nuclear Overhauser Effect Spectroscopy), to confirm the CD result. The NOESY NMR experiment measures the polarization transfer between protons (H^1), if they are close together in space, and provide all information about the distance between them. This distance information is characteristic of different secondary structures. N2A-1 was also measured using a H^1-NMR, which indicated that this protein is structured. The cross-peaks observed in the ppm range 7.6-8.6 in the NOESY spectrum demonstrate that UN2A is a highly helical protein, which is in agreement with the CD spectroscopy result. (*Fig.3.17*). The data also suggest that the protein has, at least, certain three-dimensionality, although the possibility of having an elongated, poorly compacted fold could not be excluded. Data from small-angle X-ray scattering (SAXS) would be more appropriate to conclude on the globularity of the fold.

A.



B.



C.

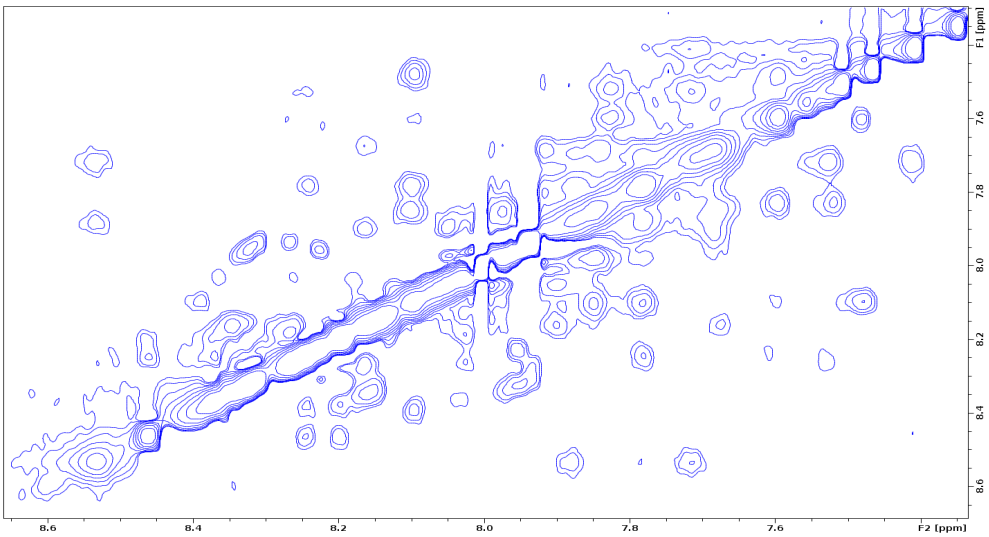


Fig.3.17 **NMR Spectra from N2A-1 and UN2A.**

A) H^1-D NMR spectroscopy of N2A-1.

B) H^1-H^1-NMR (NOESY) spectroscopy of UN2A, the red box indicates the part enlarged in **C**.

C) Enlarged view of **B**, at the ppm range 7.6-8.6. The cross-peaks observed here indicate that this protein is highly helical.

3.4 Formation and characterization of the CARP:titin N2A complex

3.4.1 CARP forms a complex with titin N2A in vitro

Previous to this study, it had been shown that CARP is able to interact with the titin N2A segment and, based on Y2H and BiYFP data, it was proposed that the minimally interacting fraction was the unique element UN2A (Miller *et al.,* 2003; Lun *et al.,* 2014). To validate complex formation and confirm the minimal binding region, we tested here binding using recombinant samples of our several titin N2A constructs and CARP$^{106\text{-}319}$ using, initially, SEC to monitor the interaction.

CARP$^{106\text{-}319}$ and N2A-1 were mixed at approximately a 3:1 molar ratio and then subjected to GF (*Fig.3.18*). The chromatogram of the mixture shows two peaks, and the elution volume of peak1 was different from the individual elution of both CARP$^{106\text{-}319}$ and N2A-1. The presence of this new peak1 may indicate complex formation (conclusion that can be reached by comparison with results from the closely related construct N2A-2; below).

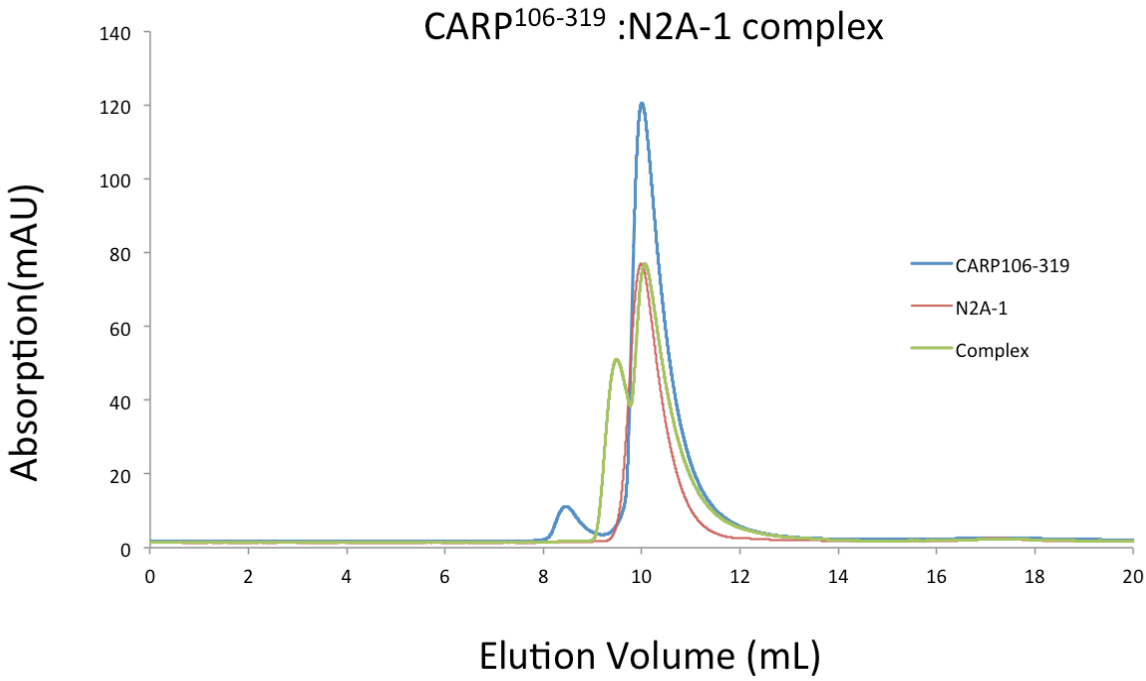


***Fig.3.18* Size exclusion chromatogram of CARP$^{106\text{-}319}$, N2A-1, and the CARP$^{106\text{-}319}$:N2A-1 complex**

Chromatograms obtained on a Superdex75 10/300 column. The curve shows A_{280} for the diverse samples.

The complexation of CARP$^{106-319}$ and N2A-2 was tested using GF at a 1:1 molar ratio mixture (*Fig.3.19*). This chromatogram shows two peaks of approximate MM 212kDa and 141kDa. On SDS-PAGE, both peaks contain almost the same amount of CARP$^{106-319}$ and N2A-2. Whether peak 1 contained the complex and/or non-interacting species that co-localized in terms of exclusion volumes was unclear. However, Peak 2 –although poorly resolved- appeared to correspond to the formed complex.

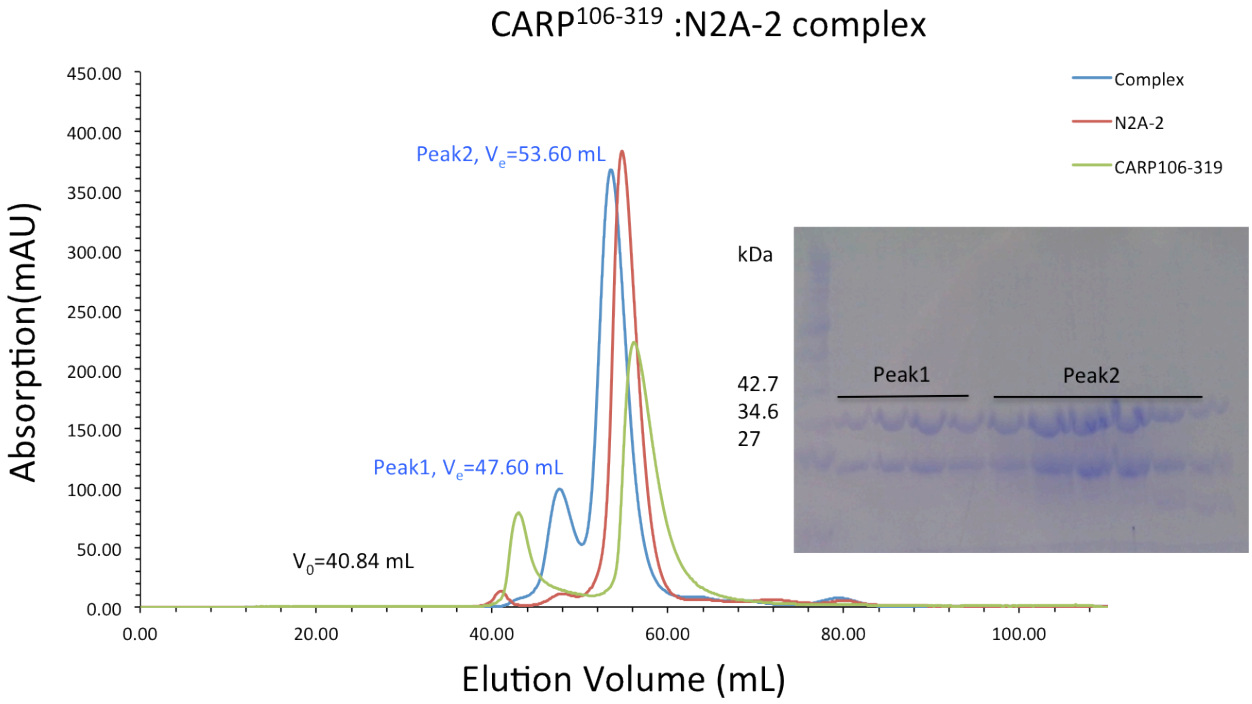


***Fig.3.19* Size exclusion chromatogram and SDS-PAGE of the complex CARP$^{106-319}$:N2A-2**
Chromatograms obtained on a Superdex75 16/60 column. The curve shows A_{280} for the diverse samples. The lanes in the SDS-PAGE correspond to the elution fractions of each peak.

Next, a 1:1 molar ratio of CARP$^{106-319}$:UN2A mixture was tested using GF. The chromatogram displays three peaks: peak1 was revealed by SDS-PAGE to correspond to apparently aggregated CARP$^{106-319}$; peak2 contained almost all CARP$^{106-319}$ sample with a small amount of co-segregated UN2A; Peak3 mostly contained UN2A and nearly no CARP$^{106-319}$. This result suggests that, unexpectedly, the binding affinity of UN2A and CARP$^{106-319}$ is not high enough to fully preserve the complex as a predominant species during GF (*Fig.3.20*).

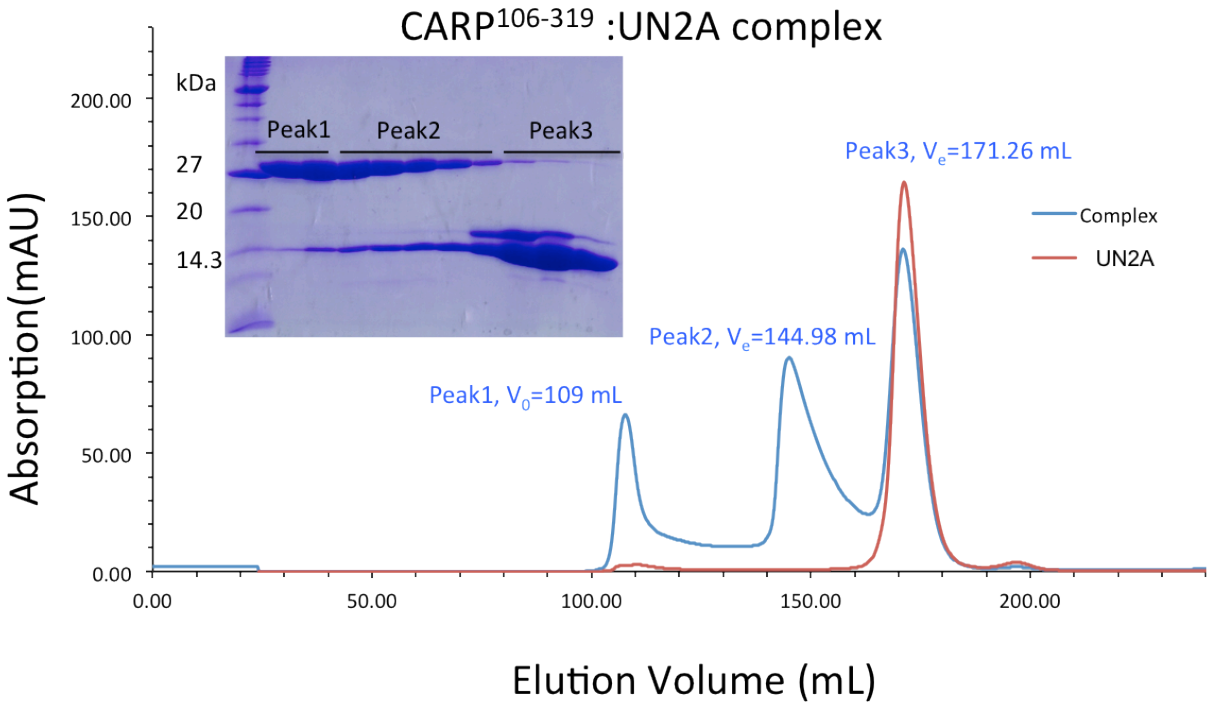


***Fig.3.20* Size exclusion chromatogram and SDS-PAGE of the CARP$^{106-319}$:UN2A complex**

Chromatograms obtained on a Superdex75 26/60 column. The curve shows A_{280} for the diverse samples. The lanes in the SDS-PAGE correspond to the elution fractions in the central part of the peak (which approximates a symmetric gaussian). The upper band of peak3 (UN2A) in the SDS-PAGE is the protein itself, with a high degree of methionine oxidation and cannot be removed by chormatographic purification.

The interaction between UN2A-Ig81 and CARP$^{106-319}$ was also tested (*Fig.3.21*). The chromatogram revealed three peaks: peak1 derived from CARP$^{106-319}$ aggregation; peak2 contains almost the same amount of CARP$^{106-319}$ and UN2A-Ig81; and peak3 corresponded to UN2A-Ig81 alone. This result suggests that these two proteins are able to interact with each other and the binding affinity is strong enough to preserve the complex during GF, so that Ig81 must be part of the CARP binding site in the titin N2A region.

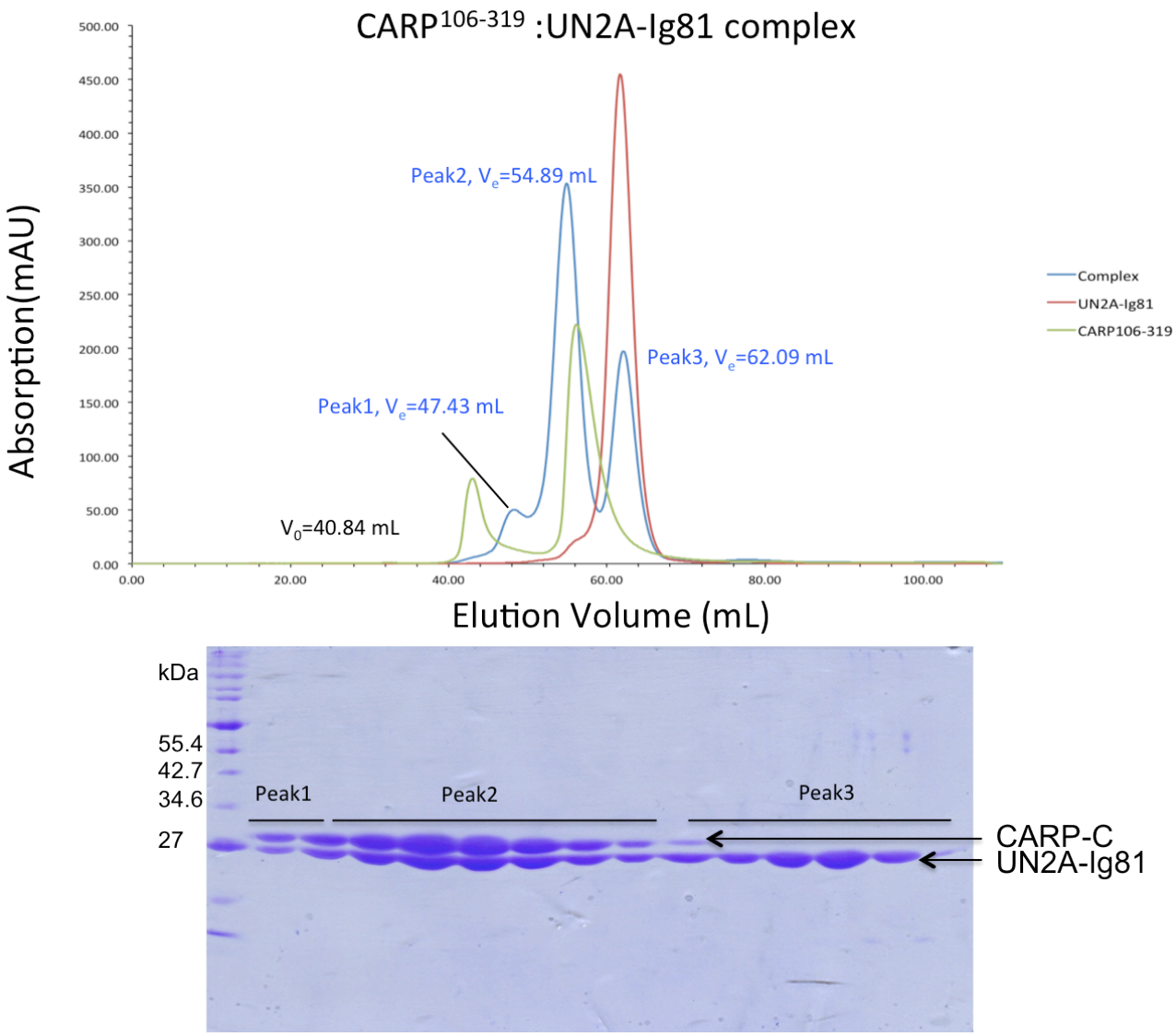


***Fig.3.21* Size exclusion chromatogram and SDS-PAGE of the CARP$^{106-319}$:UN2A-Ig81 complex**
Chromatogram obtained on a Superdex75 16/60 column. The curve shows A_{280} for the diverse samples. The lanes in the SDS-PAGE correspond to the elution fractions of the peak.

As the results indicated that CARP$^{106-319}$ is able to form a high affinity complex with UN2A-Ig81 but the interaction with UN2A has low affinity, we next tested the complexation of CARP$^{106-319}$ with the isolated Ig81 (molar ratio 1:1) to learn more about the contribution of this domain to the interaction. The resulting chromatogram showed that CARP$^{106-319}$ and Ig81 do not co-segregate (*Fig.3.22*), suggesting that this complex, if it forms, is of rather low affinity.

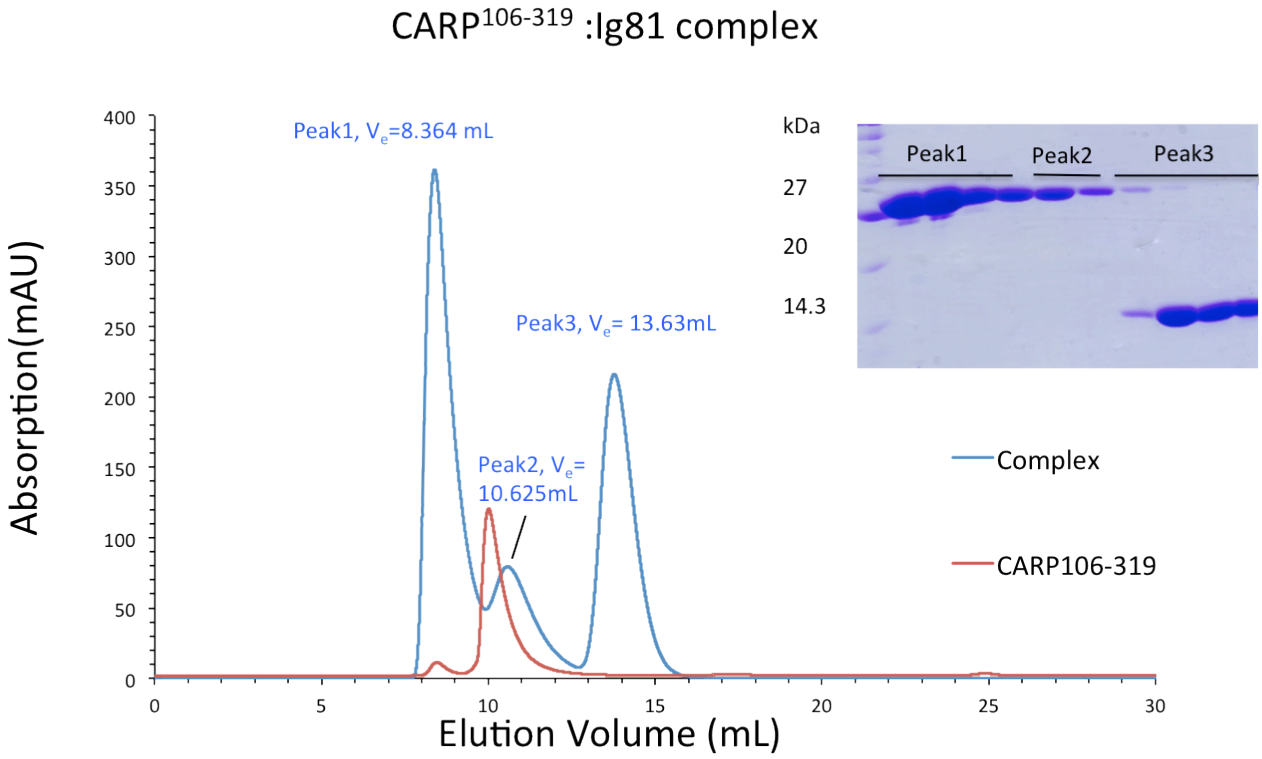


***Fig.3.22* Size exclusion chromatogram and SDS-PAGE result of the CARP$^{106-319}$: Ig81 complex**
Chromatogram obtained on a Superdex75 10/300 column. The curve shows A_{280} for the diverse samples. The lanes in the SDS-PAGE correspond to the elution fractions of the peak.

3.4.2 Estimating the Stoichiometry of the CARP:N2A complex by SEC-MALLS

The CARP$^{106-319}$:N2A-2 complex is maintained in GF and, thus, its stoichiometry could be studied using SEC-MALLS (complex was pre-purified using GF before running SEC-MALLS). This method yielded a MM of 58.44kDa for the CARP$^{106-319}$:N2A-2 complex (*Fig.3.23*). Considering that the theoretical MM of CARP$^{106-319}$ is 24.1kDa and 33.9kDa for N2A-2, a molar ratio of 1:1 of the complex can be deduced. This implies that the CARP$^{106-319}$ dimer interface is disrupted through binding to titin N2A, with the latter outcompeting the CARP counter-protomer.

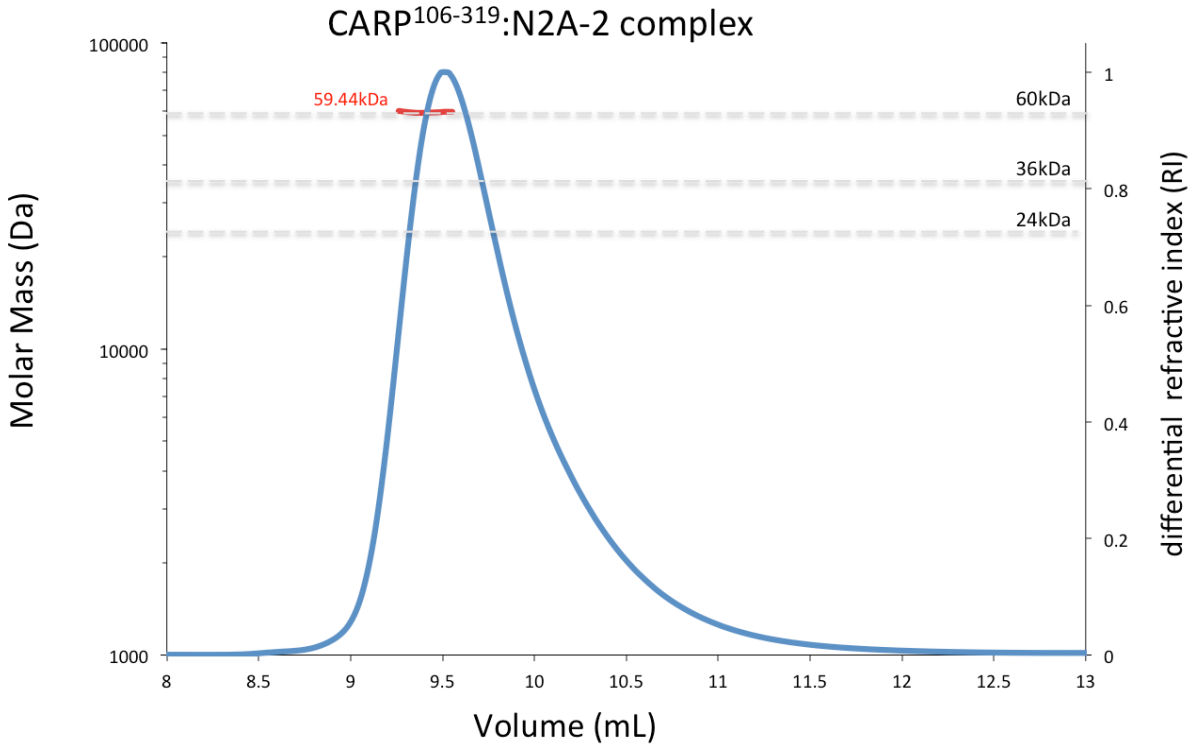


***Fig.3.23* SEC-MALLS analysis of the CARP$^{106-319}$:N2A-2 complex**

Normalized refractive index (blue) and the average MM per volume unit (red) are plotted against elution volume from a Superdex 75 10/300 size exclusion column. The theoretical MM for single species and a 1:1 complex are indicated (grey dashed lines). Complex was pre-purified using GF before running SEC-MALLS

3.4.3 Binding affinity measurement

We next aimed to measure the affinity of the CARP:UN2A-Ig81 complex . For this, we attempted measuring its K_D using microscale thermophoresis (MST), a fluorescence-based technique in solution (Wienken *et al.,* 2010). In this method, a temperature gradient is induced by an infrared laser. Protein molecules typically move out of the heated spot. This directed movement strongly depends on the properties of each molecular species (e.g. size, charge and conformation), with the change in the movement of molecules in the solution being used to calculate the K_D of binding.

In this study, the movement of molecular species was monitored through the intrinsic fluorescence of a GFP tag fused to CARP. A GFP-CARP$^{106-319}$ construct was made and purified (*Fig.3.24*). GFP-CARP$^{106-319}$:N2A-2 and GFP-CARP$^{106-319}$:UN2A-

Ig81 complexes were then formed and validated by GF and SEC-MALLS to ascertain that their characteristics reproduced those of non-GFP fusion variants. The results (*Fig.3.25; Fig.3.26; Fig.3.27; Fig.3.28*) confirmed that the GFP tag does not affect the ability of the CARP and N2A partners to interact or alter their oligomeric states.

GFP alone was also purified from GFP-CARP$^{106-319}$ construct, using AC, RAC and SEC (result not show), and served as a negative control in the MST experiment.

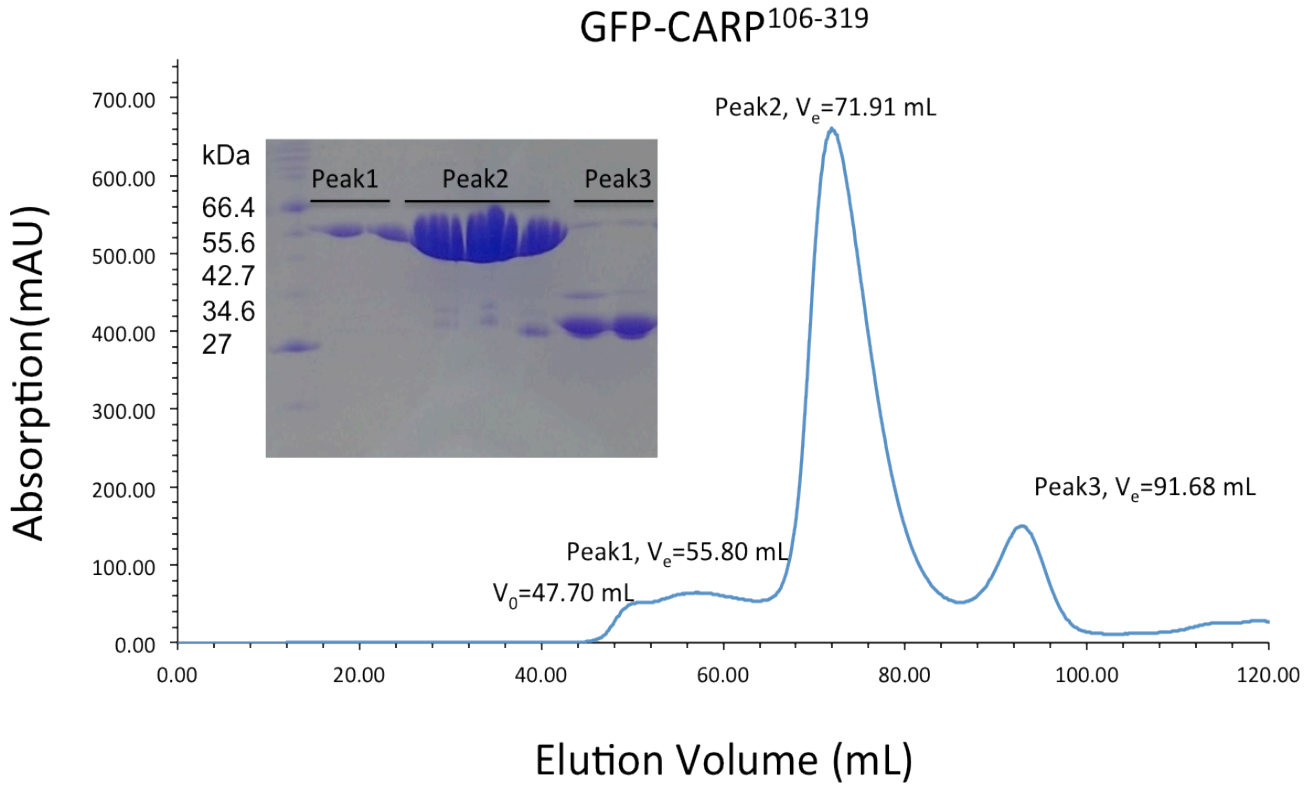


***Fig.3.24* Size exclusion chromatogram and SDS-PAGE of GFP tagged CARP$^{106-319}$**

Chromatogram (A_{280}) obtained on a Superdex200 16/60 column. The lanes in the SDS-PAGE correspond to the elution fractions in the central part of the peak.

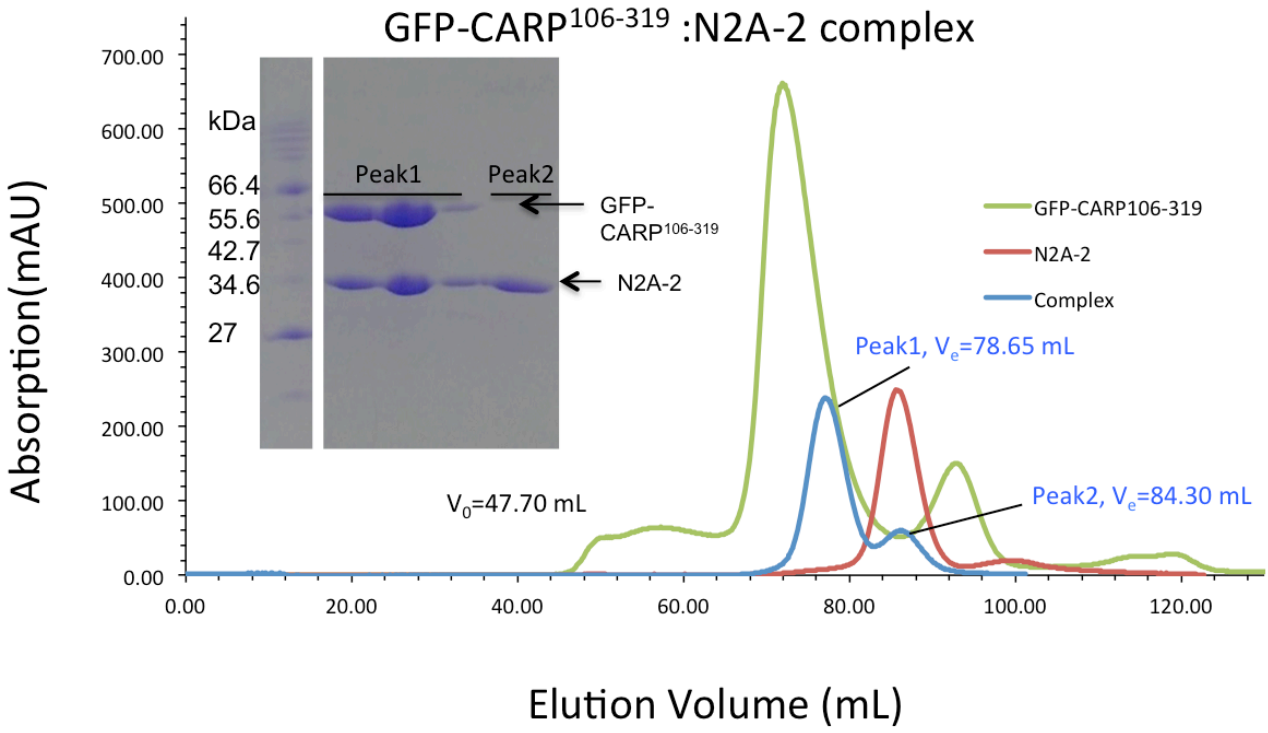


Fig.3.25 **Size exclusion chromatogram and SDS-PAGE of the GFP-CARP$^{106-319}$:N2A-2 complex**
Chromatogram obtained on a Superdex200 16/60 column. The curve shows A_{280} for the diverse samples. The lanes in the SDS-PAGE correspond to the elution fractions in the central part of the peak.

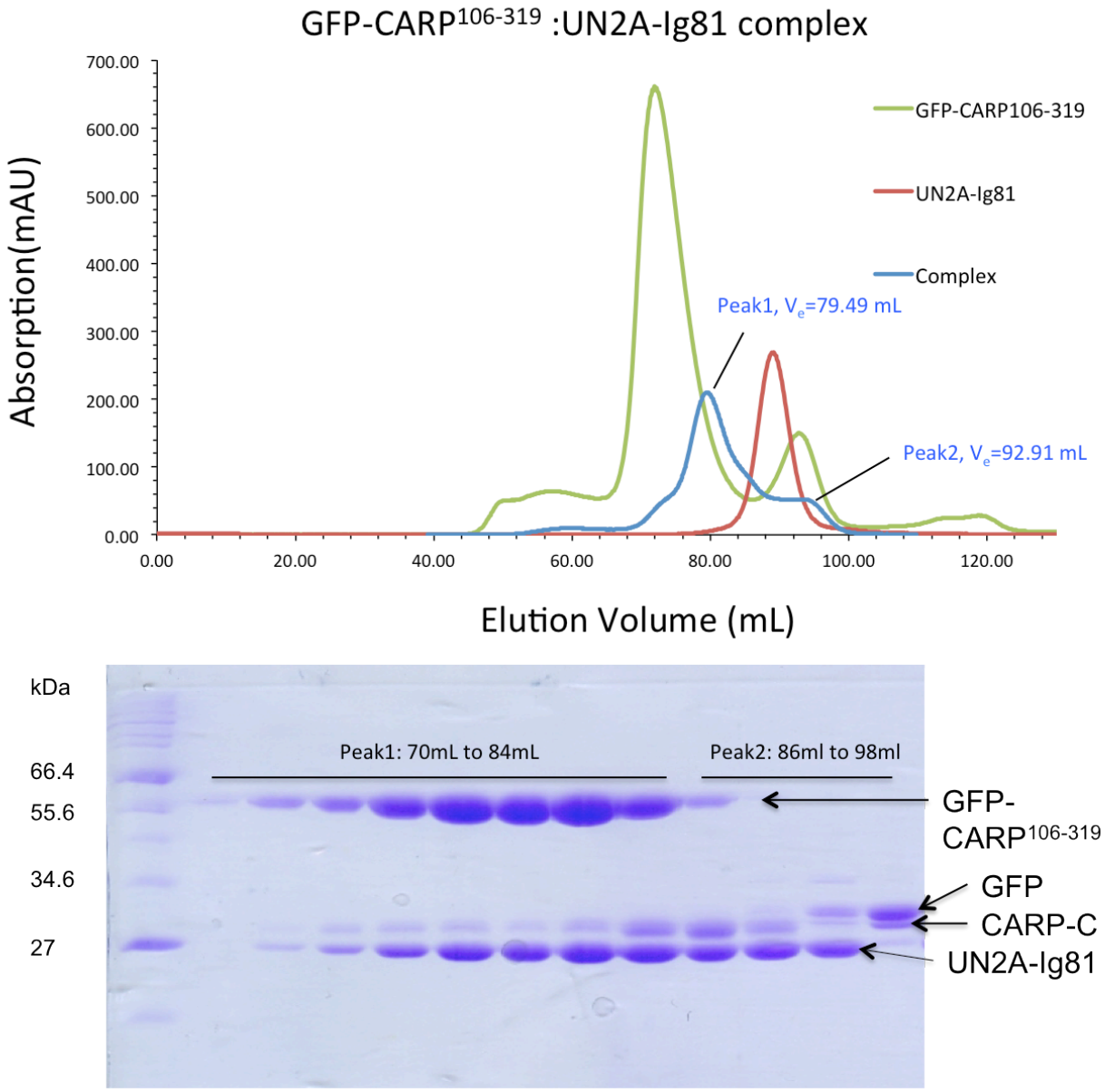


Fig.3.26 **Size exclusion chromatogram and SDS-PAGE of the GFP-CARP$^{106-319}$:UN2A-Ig81 complex**

Chromatogram obtained on a Superdex200 26/60 column. The curve shows A_{280} for the diverse samples. The lines in the SDS-PAGE correspond to the whole chromatograph (2mL per each band). A certain amount of degradation of GFP-CARP$^{106-319}$ (approximately at the fusion point) was revealed.

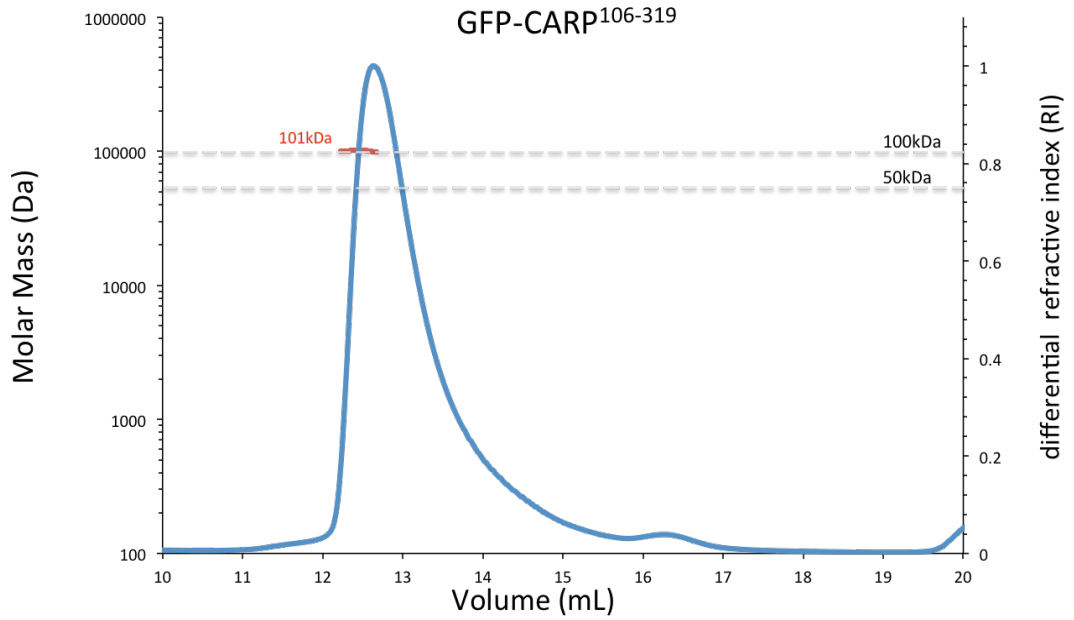


Fig.3.27 **SEC-MALLS analysis of GFP-CARP$^{106-319}$**

Normalized refractive index (blue) and measured MM (red) are plotted against the elution volume using a Superdex75 10/300 size exclusion column. The average MM measured was 100kDa. Theoretical MM for monomer and dimers of this sample are indicated (50kDa). Complex was pre-purified using GF before running SEC-MALLS

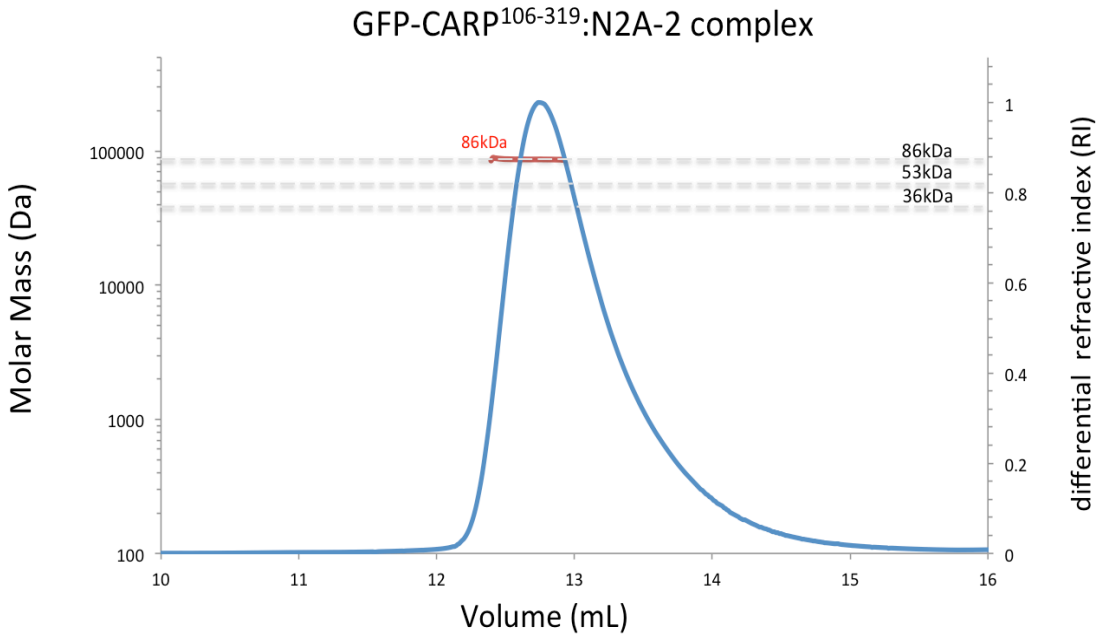


Fig.3.28 **SEC-MALLS analysis of GFP-CARP$^{106-319}$:N2A-2 complex**

Normalized refractive index (blue) and measured MM (red) are plotted against elution volume from a Superdex 75 10/300 size exclusion column. The theoretical MM for single species and a 1:1 complex are indicated (grey dashed lines). Complex was pre-purified using GF before running SEC-MALLS

The proteins assayed in MST were in GF buffer with 0.05% Tween and 0.3mg BSA, the capillary used was Premium coated. Different GFP-CARP$^{106-319}$ concentrations were tried (80nM, 40nM and 20nM) with the concentration of N2A-1 diluted 10 times per each capillary tested (14 total in the titration series), from 270μM to 27x10^{-12}nM. Unfortunately, the MST experiments did not yield interpretable results pointing for the need of optimization in the used protocol.

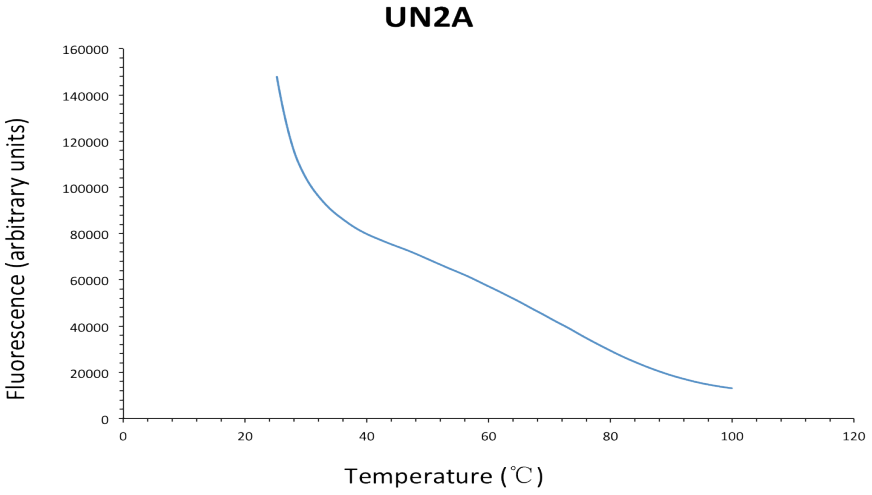
Given our limited access to the MST instrumentation, further MST experimentation became impractical and alternative methods had to be evaluated. Isothermal titration calorimetry (ITC) was considered. However, because CARP$^{106-319}$ is a dimer and the complex is formed in a 1:1 ratio, the interaction has two sources of heat (dimeric CARP breaking and the complex with N2A forming). This highly complicates the interpretation of ITC data, troubling the calculation of a reliable K_D. As an alternative, we considered using Differential Scanning Fluorimetry (DSF). DSF is a fluorescence-based technique that is used to measure the Tm of proteins (Frank *et al.,* 2007). With increasing temperature, a protein loses its structure and its hydrophobic core becomes exposed. The Sypro-orange dye then interacts with the accessible hydrophobic groups and emits fluorescence, so that the Tm can be calculated from the change in fluorescence intensity. If the Tm of two target proteins can be measured using this method and values are sufficiently different as to give resolved curves, the K_D of the interaction, in theory, can also be obtained by measuring the shift of the Tm value of the thermolabile protein upon titration with the binding partner. This method was deemed to be particularly suitable for the current study as the UN2A construct presents an unusually high thermal stability and does not yield a melting curve (*Fig.3.29*).

Therefore, the Tm values of four proteins (*Table.3.2*) were measured using DSF and, for each protein, four different concentrations were tested (2μM, 5μM, 10μM, 40μM). The DSF result of CARP$^{106-319}$ and Ig81 shows their Tm to be 43°C for CARP$^{106-319}$ (in excellent agreement with values calculated using CD) and 67°C for Ig81 (*Fig.3.30*, Tm calculation result not shown). However, the UN2A and UN2A-Ig81 samples produced a very high fluorescence signal at the beginning of the experiment, obscuring any potential monitoring of fluorescence changes in the CARP counterpart.

Table.3.2 Conditions of DSF assays. Each mixture is aliquoted into three (10 μL each) for the measurement.

Final protein concentration	Protein sample (50 μM)	Dye (50x)	GF buffer	Total
2 μM	1.4 μL	3.5 μL	30.1 μL	35 μL
5 μM	3.5 μL	3.5 μL	28 μL	35 μL
10 μM	7 μL	3.5 μL	24.5 μL	35 μL
40 μM	28 μL	3.5 μL	3.5 μL	35 μL

A.



B.

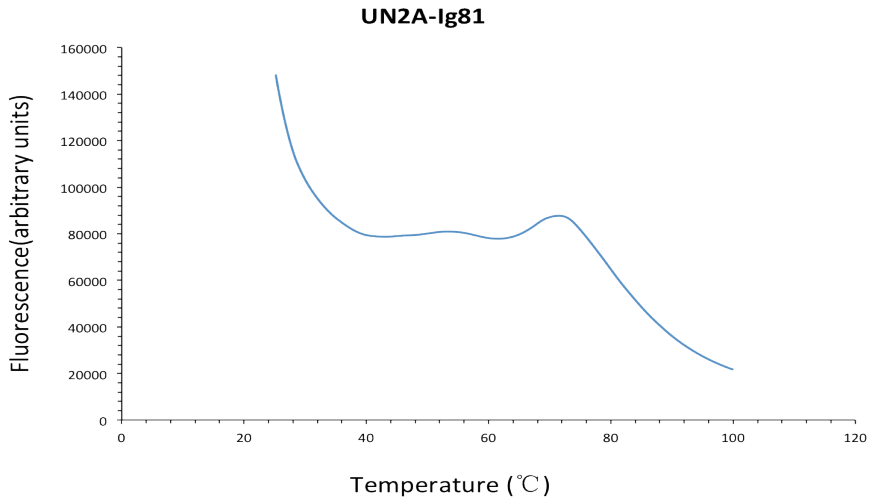
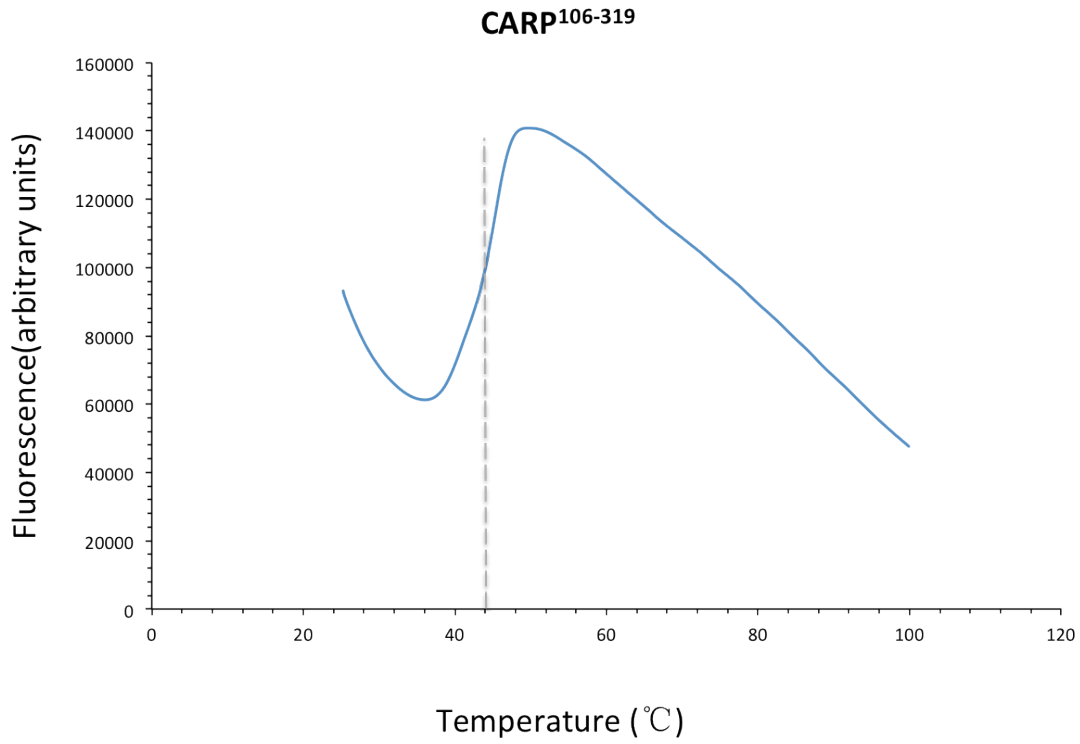


Fig.3.29 **The DSF result of UN2A and UN2A-Ig81**

A) The curve shows that UN2A start with a high signal of fluorescence.

B) The curve shows that UN2A-Ig81 start with a high fluorescence signal.

A.



B.

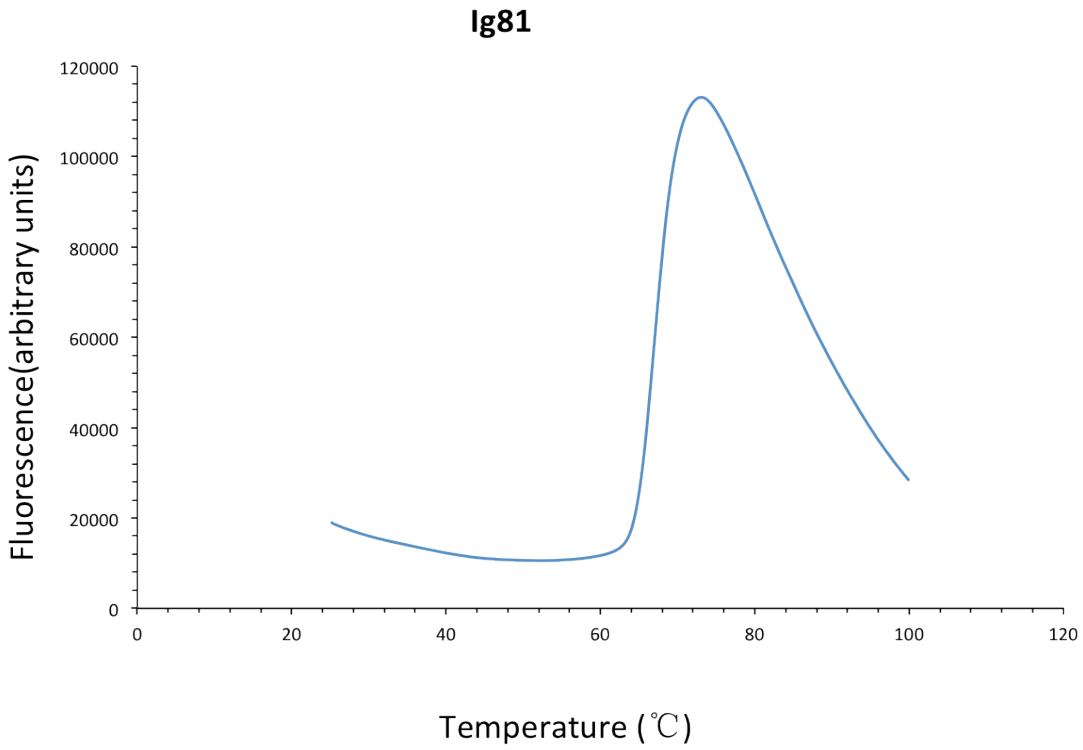


Fig.3.30 **The DSF result of CARP$^{106-319}$ and Ig81**

A) DSF curve of CARP$^{106-319}$ shows that this protein has a Tm=43°C.

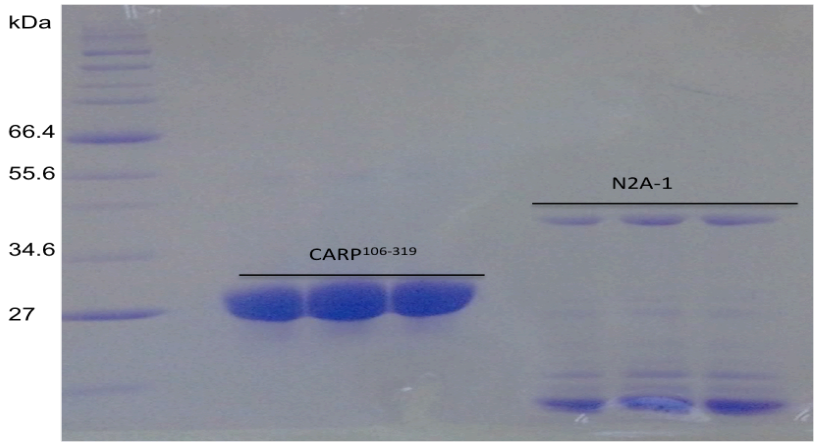
B) DSF curve of Ig81 shows that the Tm of this protein is 67°C.

3.5 Atomic characterization of CARP, titin N2A and their complex

3.5.1 Stability of CARP and titin N2A samples

The stability of CARP$^{106-319}$ and N2A samples upon storage (4°C for 14 days) was monitored using SDS-PAGE. CARP$^{106-319}$ appeared stable during this time. However, N2A-1 underwent strong degradation (*Fig.3.31*). In N2A-2, the degradation was far less prominent. Unfortunately, N2A-2 degradation could not be prevented by complexation with CARP$^{106-319}$. But, the small levels of degradation suggested that both isolated N2A-2 and its CARP$^{106-319}$-complexed form are suitable for structural and biophysical study.

A.



B.

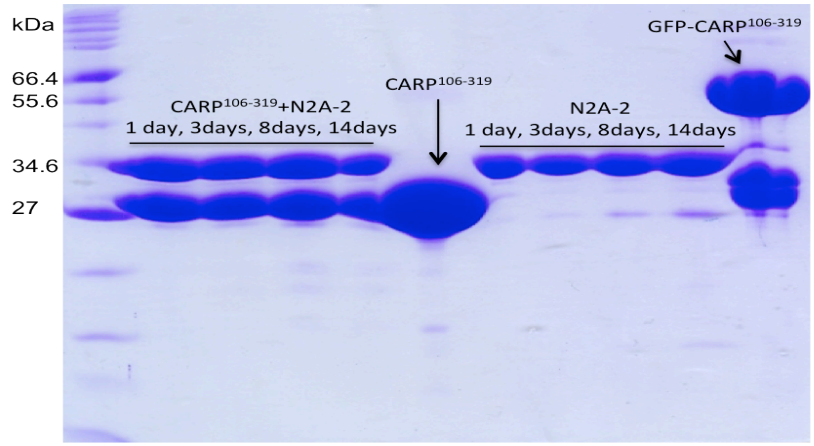


***Fig.3.31* Stability check of CARP and N2A samples by SDS-PAGE**

A) CARP$^{106-319}$ and the N2A-1 samples assayed on SDS-PAGE. Each samples has been tested in triplicate after 14 days in 4°C (replicates).

B) CARP$^{106-319}$, GFP- CARP$^{106-319}$, N2A-2, and the CARP$^{106-319}$:N2A-2 mixture, on SDS-PAGE. The mixture and N2A-2 was measured four times, with a different time scale: one day, three days, eight days and fourteen days.

3.5.2 Crystallization of CARP$^{106-319}$ and titin N2A proteins

In order to characterize the CARP/N2A interaction structurally at atomic level, we pursued the crystallization of isolated and complexed samples. CARP$^{106-319}$; Ig81; UN2A and the CARP$^{106-319}$:UN2A-Ig81 complex were assayed.

CARP$^{106-319}$ was trialed at both room temperature and 4°C. The protein was in GF buffer (supplemented with 3mM β-ME, *Table.3.4*) at several initial concentrations: 60mg/ml, 30mg/ml, 20mg/ml and 10mg/ml (*Table.3.4*). Initial crystals of CARP$^{106-319}$ were observed in condition A1 of the NATRIX screen (Hampton Research) containing 0.05M MES pH5.8, 1.8M Li_2SO_4 and 0.01mM $MgCl_2$ incubated at 19 °C (*Fig.3.32*).

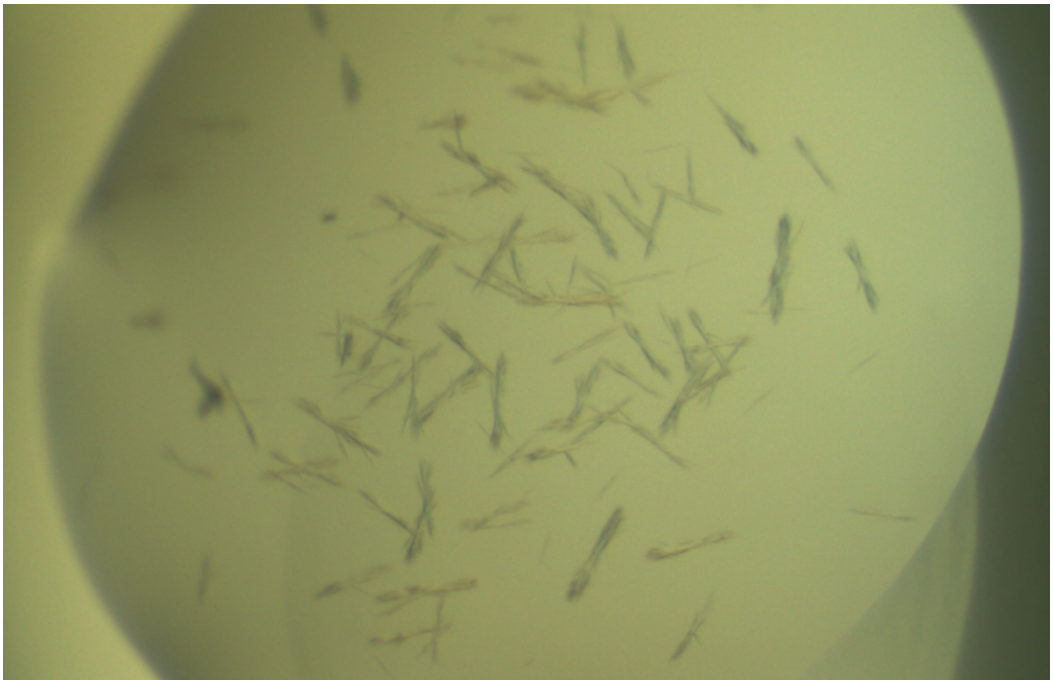


Fig.3.32 **Crystals grown from CARP$^{106-319}$**

Needle like protein crystals were observed. The condition is 0.05M MES pH5.8, 1.8M Li_2SO_4 and 0.01mM $MgCl_2$. The precipitant:protein ratio (nL) is 100:100. The length of the crystals is approximate 200μm.

Based on the initial NATRIX-A1 condition, several optimization screens were designed (OPT1-OPT6) (*Fig.3.33*). However, crystals were only grown in 2 drops from one of the screens (OPT2)(*Fig.3.34*), with all subsequent attempts (OPT3-OPT6) failing to yield any crystalline samples. We concluded that these crystals

are of very low reproducibility and might not lead to a viable crystallographic structure elucidation.

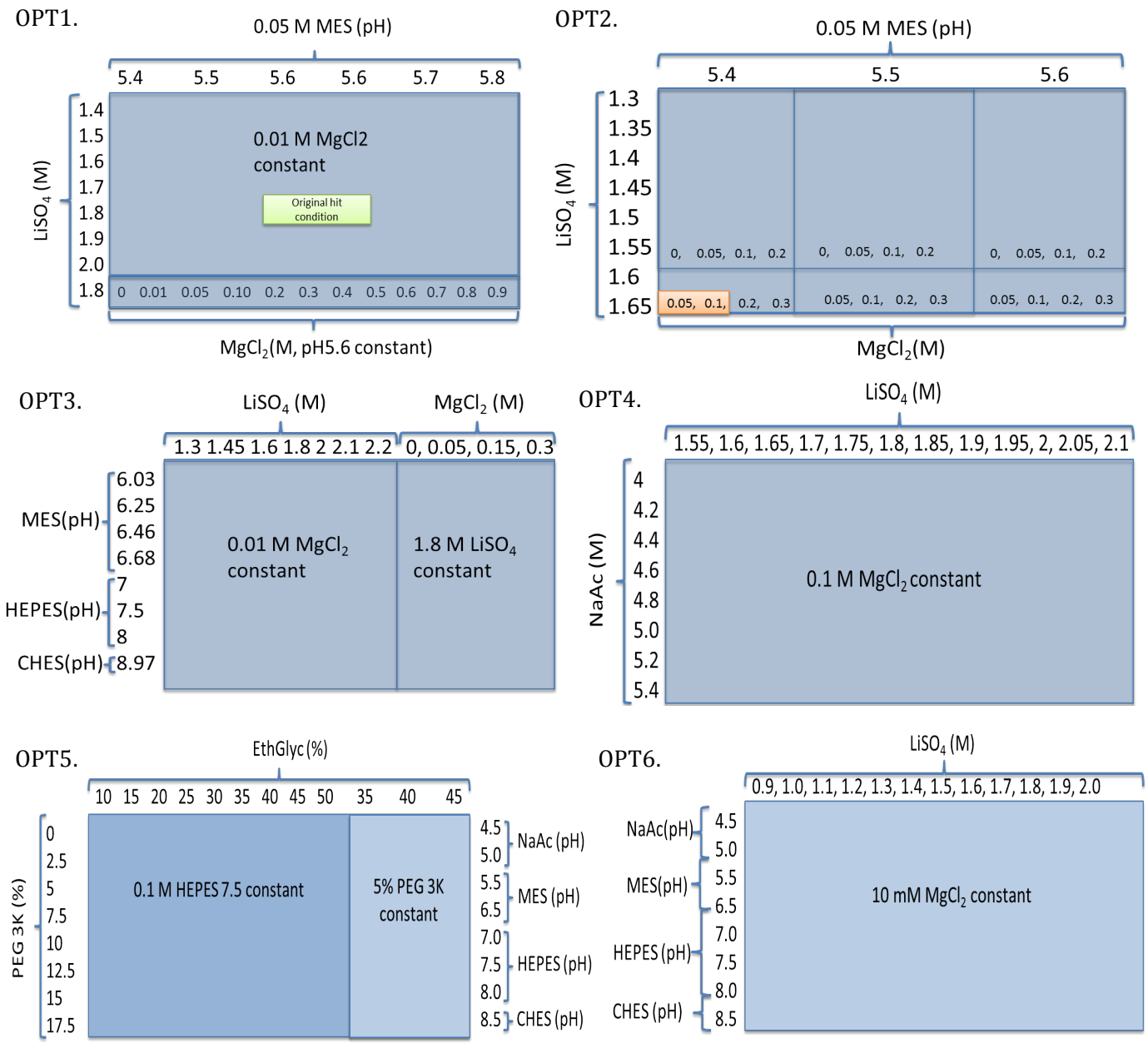


Fig.3.33 **Optimization screens of CARP$^{106-319}$**

All of the optimization screens of CARP$^{106-319}$ (OPT1-OPT6)

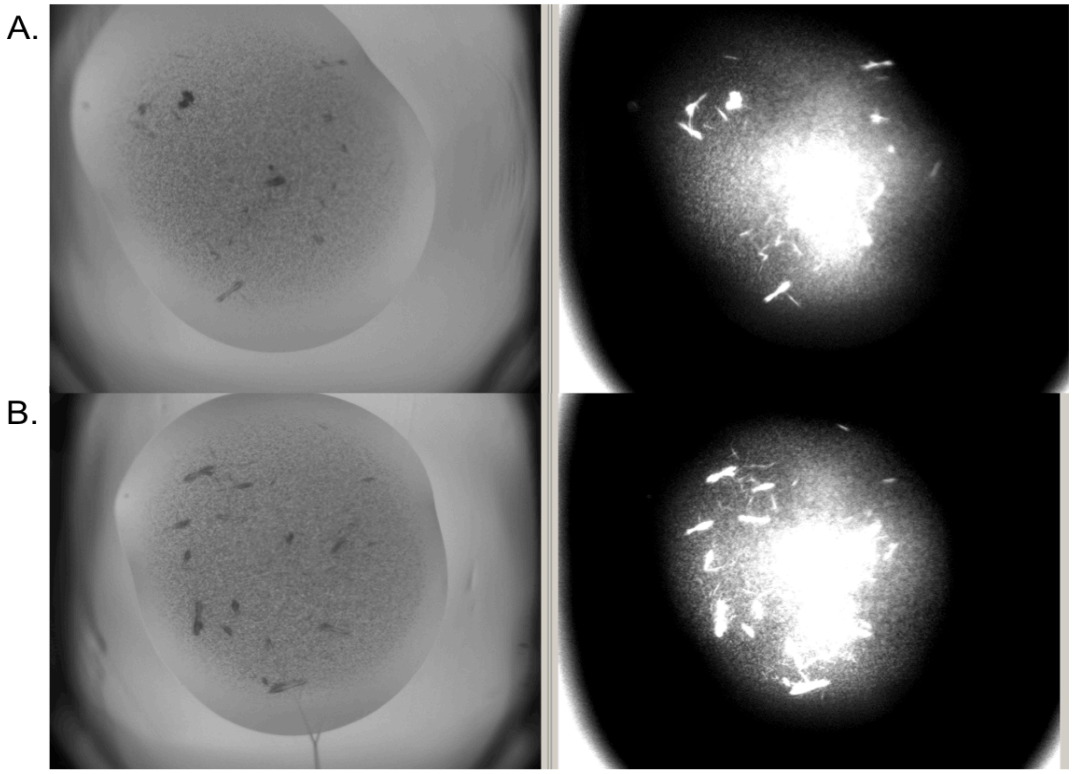


Fig.3.34 **Crystals of CARP$^{106-319}$ observed in OPT2**

A) Crystals observed in 0.05M MES pH5.8, 1.65M Li_2SO_4 and 0.05mM $MgCl_2$. The precipitant:protein ratio (nL) is 100:100. The length of the crystals is approximate 100μm.

B) Crystals observed in 0.05M MES pH5.8, 1.65M Li_2SO_4 and 0.1mM $MgCl_2$. The precipitant:protein ratio (nL) is 100:100. In both cases, UV illumination proves their protein composition. The length of the crystals is approximate 100μm.

UN2A was trialed at both room temperature and 4°C. The protein was in GF buffer with initial concentrations of 30mg/ml and 78mg/ml (*Table.3.3*). To date, no crystals of this sample have been observed. However, the quality of NMR spectra obtained to date suggests that the 3D-structure of UN2A could be obtained this way. To that effect, the production of UN2A using minimum media (without isotope labeling) was tested and over 30mg of pure protein obtained from 1L culture. This suggests a viable NMR structural determination, to be considered in future work.

The purified CARP$^{106-319}$:UN2A-Ig81 complex was trialed at room temperature. The complex was in GF buffer at an initial concentration of 30mg/ml (*Table.3.4*). Several small, multiple crystals were observed, where UV illumination demonstrated them to contain protein. However, the quality of these crystals was not high enough to permit conducting X-ray diffraction tests (*Fig.3.35; Fig.3.36*).

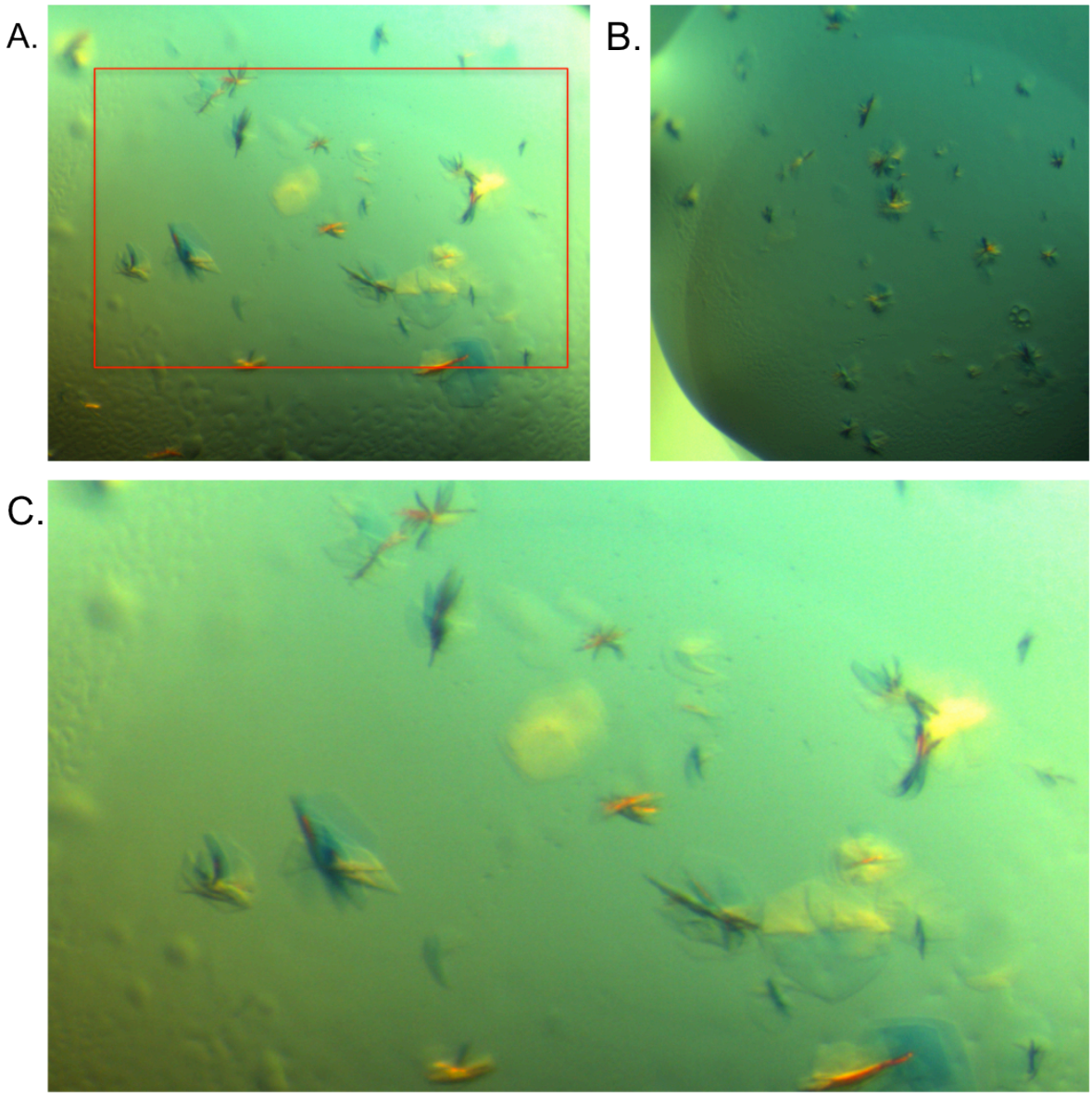


Fig.3.35 **Stacks of crystal plates of the CARP$^{106-319}$:UN2A-Ig81 complex.**

A) Crystals observed in MORPHEUS (Molecular Dimension) B12A. The condition is 0.09M Halogens (0.3M Sodium fluoride; 0.3M Sodium bromide; 0.3M Sodium iodide), 0.1M Tris-BICINE pH8.5, 50% v/v precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350). The precipitant:protein ratio (nL) is 200:200. The red box area is enlarged in **C**. The length of the crystals is approximate 150µm.

B) Crystals observed in MORPHEUS (Molecular Dimension) D12A. The condition is 0.12M Alcohol (0.2M 1,6-Hexanediol; 0.2M 1-Butanol; 0.2M 1,2-Propanediol; 0.2M 2- Propanol; 0.2M 1,4-Butanediol; 0.2M 1,3-Propanediol), 0.1M Tris-BICINE pH8.5, 50% v/v precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350). The precipitant:protein ratio (nL) is 200:200. The length of the crystals is approximate 100µm.

C) The enlarged view of **A** (in red box).

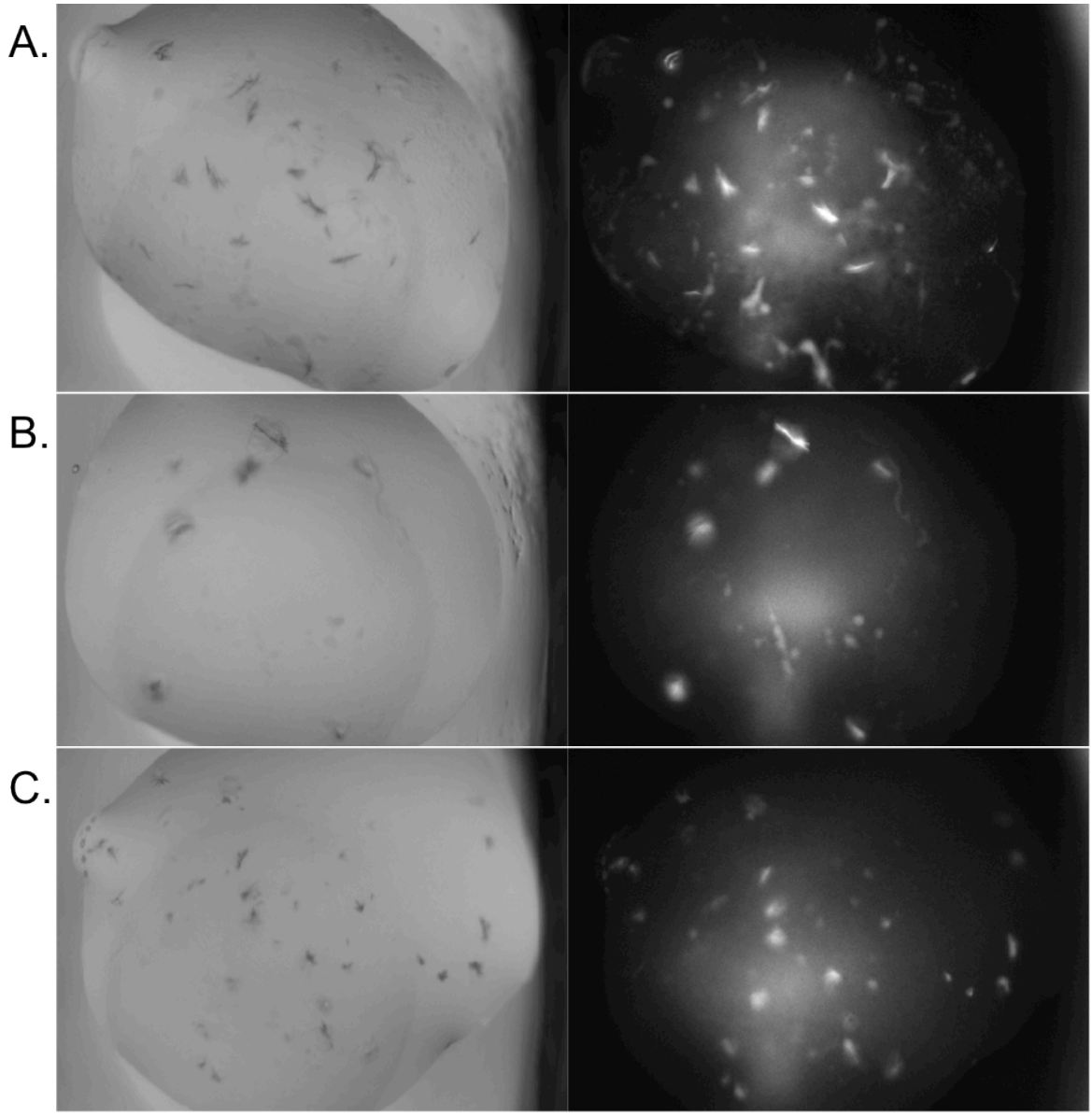


Fig.3.36 **UV imaging of CARP$^{106-319}$:UN2A-Ig81 crystals**

A) UV imaging of crystals grown in MORPHEUS (Molecular Dimension), condition B12A: 0.09M Halogens (0.3M Sodium fluoride; 0.3M Sodium bromide; 0.3M Sodium iodide), 0.1M Tris-BICINE pH8.5, 50% v/v precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350). The precipitant:protein ratio (nL) is 200:200. The length of the crystals is approximate 150µm.

B) UV imaging of new crystals obtained from MORPHEUS (Molecular Dimension) C12A. The condition is 0.09M NPS (0.3M Sodium nitrate, 0.3 Sodium phosphate dibasic, 0.3M Ammonium sulfate), 0.1M Tris-BICINE pH8.5, 50% v/v precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350). The precipitant:protein ratio (nL) is 200:200. The length of the crystals is approximate 100µm.

C) UV imaging of crystals grown from MORPHEUS (Molecular Dimension) D12A. The condition is 0.12M Alcohol (0.2M 1,6-Hexanediol; 0.2M 1-Butanol; 0.2M 1,2-Propanediol; 0.2M 2- Propanol; 0.2M 1,4-Butanediol; 0.2M 1,3-Propanediol), 0.1M Tris-BICINE pH8.5, 50% v/v precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350). The precipitant:protein ratio (nL) is 200:200. The length of the crystals is approximate 100µm.

Additionally, the Ig81 domain in isolation was also trialed at room temperature. The protein was in GF buffer with initial concentrations of 15mg/ml and 38.5mg/ml (*Table.3.4*). Several crystals were obtained from the higher concentration. Most crystals were observed after two weeks, but they were too small and fragile for mounting for X-ray diffraction (*Fig.3.37*). However, after two months, the two best crystals could be harvested and their diffraction properties tested at the Diamond synchrotron (*Fig.3.38,* crystals in **B** were used in diffraction tests). Crystals were prepared for cryo-crystallography by vitrification at 100k in cryo-protection medium (20% ethylene glycol; 20% isoproanol; 0.2M $MgCl_2$; 0.1M Tris pH 8.5) and mounted on a 0.1mm litho loop (Molecular Dimension). They were vitrified into a pre-cooled puck and shipped to the synchrotron in their frozen state. Diffraction data were recorded to 2.15 Å resolution. The lattice was identified as $P6_{(X)}22$ with unit cell dimensions a=b=115.312Å, c=52.226Å (α=β=90°, γ=120°). An X-ray diffraction pattern is shown in *Fig.3.39* and the data processing statistics are presented in *Table.3.3*.

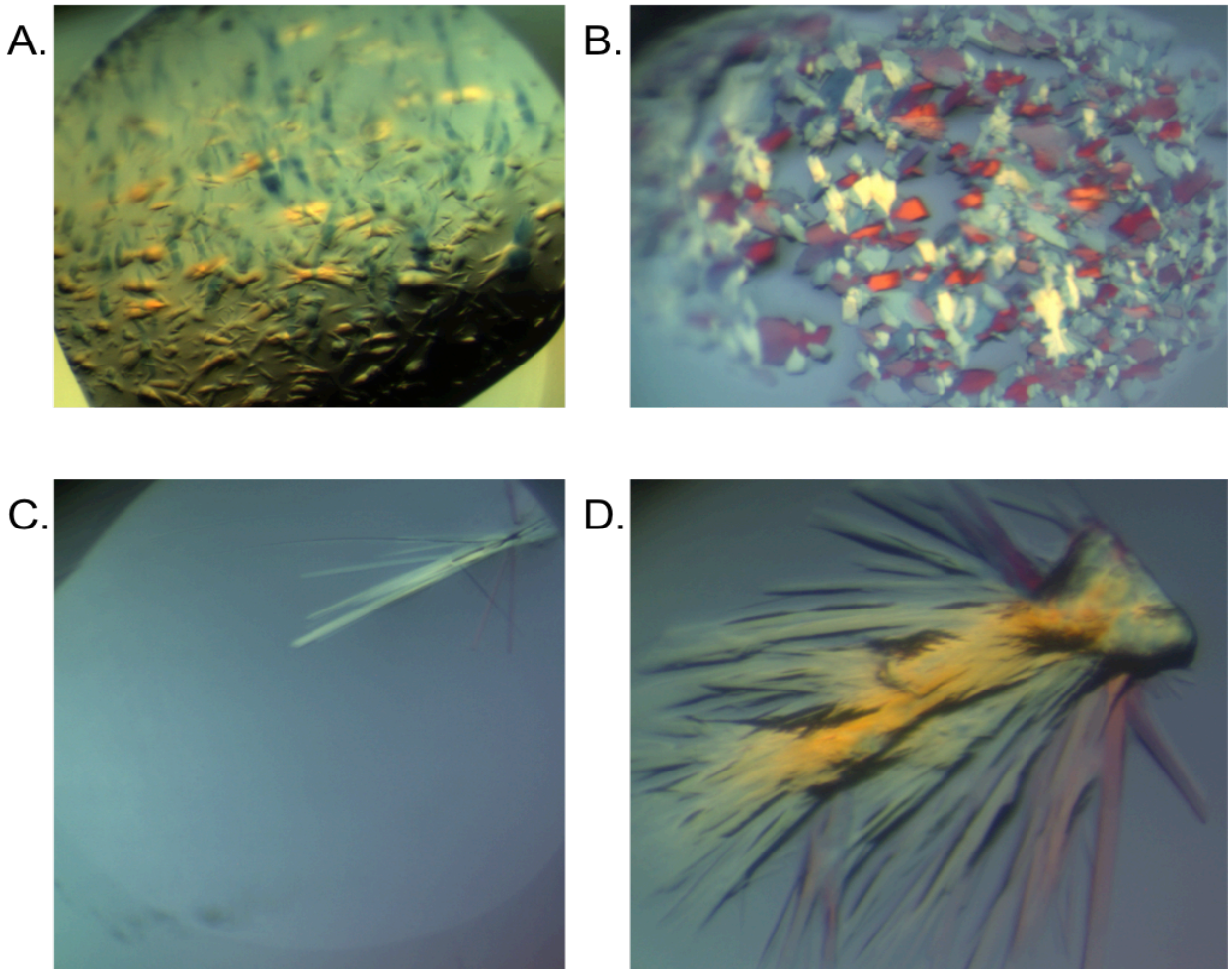


Fig.3.37 **Crystals of Ig81 grown in two weeks**

A) Crystals grown in NATRIX. Mother liquor condition: 0.01M $Mg(CH_3COO)_2$, 0.05M MES pH5.6, 2.5M $(NH_2)_2SO_4$. The precipitant:protein ratio (nL) is 200:200.

B) Crystals grown from Structure screen. Mother liquor condition: 2M $MgCl_2$, 0.1M BICINE pH9.0. The precipitant:protein ratio (nL) is 100:200.

C) Needle like crystals observed in Structure screen 1+2, grown from 40% (v/v) PEG300, 100mM HEPES/NaOH pH7.5, 200mM NaCl. The precipitant:protein ratio (nL) is 100:200.

D) Crystal in C after 5 days.

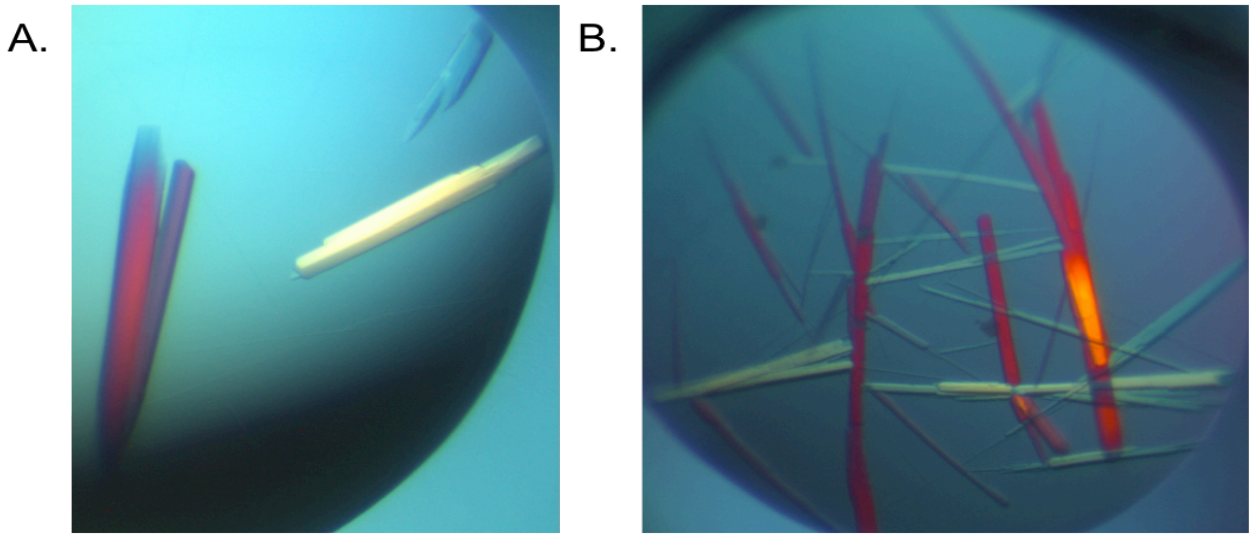


Fig.3.38 **Rod-like crystals of Ig81 grown for two months**

A) The condition is 0.2M magnesium chloride hexahydrate; 0.1M Tris pH 8.5; 30% w/v PEG 4000. The precipitant:protein ratio (nL) is 200:200.

B) Same condition as **A**, The precipitant:protein ratio (nL) is 200:100.

Table.3.3: X-ray diffraction data processing statistics calculated using XSCALE (Kabsch W, 2010)

Wavelength	0.9282Å
Resolution	30.00Å -2.15Å (2.2Å -2.15Å)
R merge	8.3% (52.4%)
Unique reflections	11592 (740)
Multiplicity	11.4 (11.9)
Completeness	99.8% (100%)
CC ½	0.999 (0.951)
I/ σ (I)	17.4 (5.23)

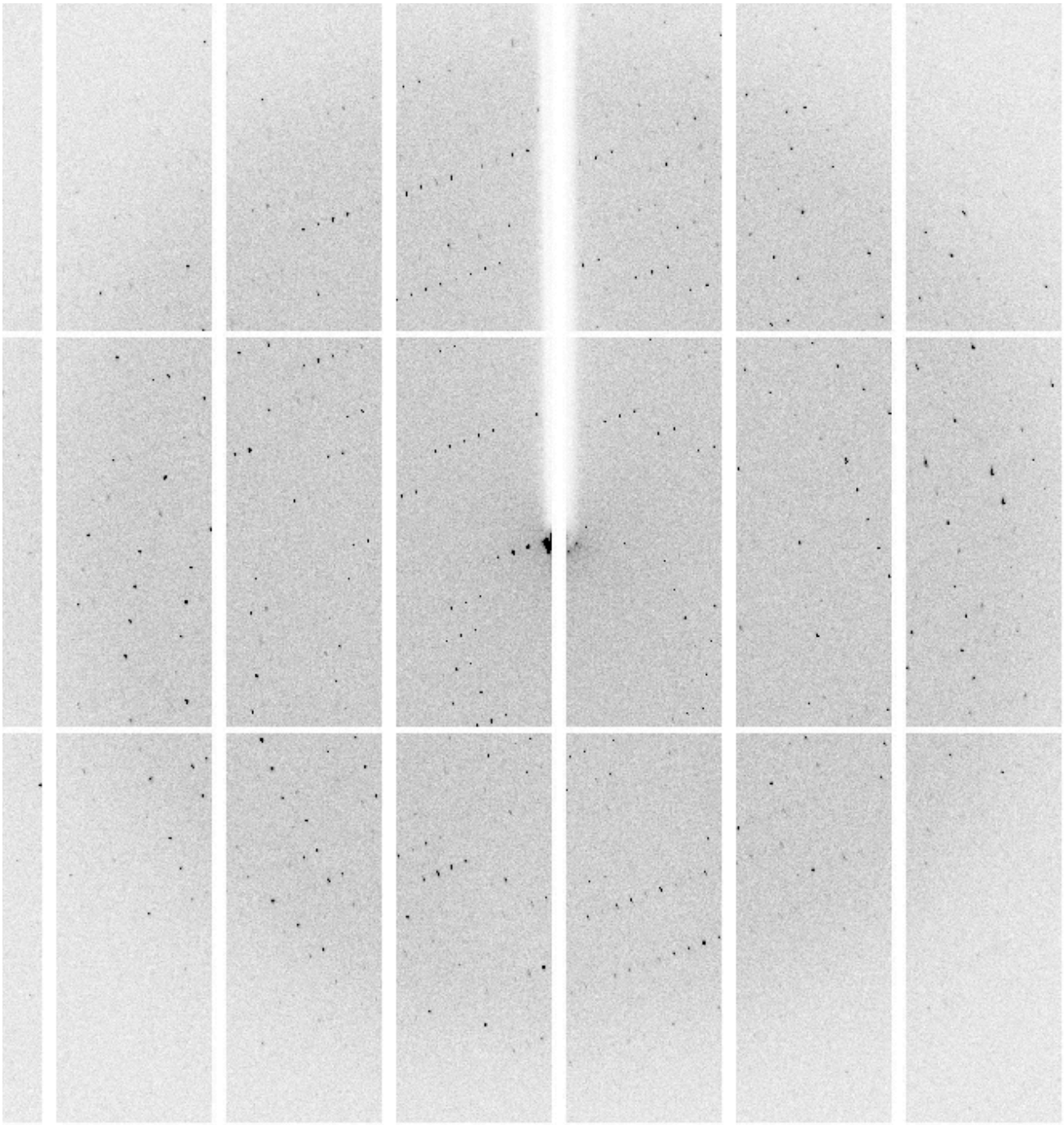


Fig.3.39 The X-ray diffraction pattern of Ig81 crystals

X-ray diffraction image from Ig81 crystals shown in *Fig 3.37B* . The isotropic resolution at the edge of the detector is 2Å.

All trialed proteins and the screens used are shown in *Table.3.3*.

Table.3.4 Trialed proteins and the corresponding screens. Screens that yielded crystals are given in red.

Protein	Buffer condition	Protein concentration	Size (nL, precipitant: protein)	Screen condition
CARP$^{106-319}$	25mM HEPES, 100mM NaCl, 3mM β ME	A: 60mg/ml B: 30mg/ml	100:100	JCSG, PEGRx, CRYO, WIZARD, MORPHEUS, PACT-Premier
CARP$^{106-319}$	25mM HEPES, 100mM NaCl	A: 20mg/ml	100:100	**RT:** Stru SC1+2, SALTRX, NATRIX (A1), WIZARD **4°C:** Stru SC1+2, SALTRX, NATRIX, WIZARD
CARP$^{106-319}$	25mM HEPES, 100mM NaCl	A: 20mg/ml	100:100	OPT1, PACT, PEGRx, CRYO (C11), MORPHEUS, WIZARD
CARP$^{106-319}$	25mM HEPES, 100mM NaCl	A: 20mg/ml B: 20mg/ml	200:200	OPT1, OPT2
CARP$^{106-319}$	25mM HEPES, 100mM NaCl	A: 20mg/ml B: 20mg/ml	100:100	OPT3
CARP$^{106-319}$	25mM HEPES, 100mM NaCl	A: 20mg/ml B: 10mg/ml	100:100	OPT4
CARP$^{106-319}$	25mM HEPES, 100mM NaCl	A: 20mg/ml B: 20mg/ml	100:100	OPT5, OPT6
UN2A	25mM HEPES, 100mM NaCl	A: 30mg/ml B: 15mg/ml	100:100	CRYO, NATRIX, Stru Sc1+2, WIZARD, PEGRx, MORPHEUS
UN2A	25mM HEPES, 100mM NaCl	B: 78mg/ml	200:200	**RT:** CRYO, NATRIX, Stru Sc1+2, WIZARD, PEGRx, MORPHEUS, SALTRX **4℃:** CRYO, NATRIX, PEGRx, MORPHEUS, SALTRX
Ig81	25mM HEPES, 100mM NaCl	A: 15mg/ml B: 15mg/ml	A: 200:200 B: 200:100	CRYO, PACT, MORPHEUS, PEGRx, WIZARD,NATRIX
Ig81	25mM HEPES, 100mM NaCl	A: 38.5mg/ml B: 38.5mg/ml	A: 200:200 B: 100:200	CRYO (A4B; A5B), PACT, MORPHEUS, Stru Sc 1+2 (C9A; C9B; E2B; E7B; F6B;), JCSG, NATRIX (A2A; B7B;F12A)
CARP$^{106-319}$:UN2A-I81	25mM HEPES, 100mM NaCl	A: 30mg/ml B: 30mg/ml	A: 200:200 B: 200:100	CRYO, SALTRx, MORPHEUS (B12A; C12A; D12A), PEGRx, WIZARD, NATRIX

3.6 Summary of results

A summary of all the results acquired in this work is shown in *Table.3.5*.

Table.3.5 Summary of results obtained.

Constructs	MM (kDa)	Solubility	Yield (per 1L)	Stability (14 days)	Oligomeric state	Tm	Crystallization trays	Initial Crystals	Interaction affinity with CARP
$CARP^{106-319}$	24.1	Soluble	7mg	Stable	Dimer	43°C	Set	Yes	N/A
GFP-$CARP^{106-319}$	53	Soluble	30mg	Unstable	Dimer	N/A	N/A	N/A	N/A
N2A-1	36	Soluble	8mg	Unstable	Monomer	N/A	N/A	N/A	High
N2A-2	35	Soluble	8mg	Unstable	Monomer	N/A	N/A	N/A	High
UN2A	13.3	Soluble	30mg	Stable	Monomer	N/A	Set	N/A	Low
Ig80	10	Insoluble	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ig80-UN2A	24	Insoluble	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ig81	10.3	Soluble	8mg	Stable	Monomer (Not measured by SEC-MALLS)	67°C	Set	Yes	Low
UN2A-Ig81	23.6	Soluble	10mg	Stable	Monomer (Not measured by SEC-MALLS)	N/A	Set (with $CARP^{106-319}$)	Yes (with $CARP^{106-319}$)	High

4. Discussion

4.1 Biophysical characterization of proteins

4.1.1 The C-terminal AR-domain of CARP forms weak dimers

The N-terminal region (residues 1-122) of CARP was previously thought essential to the formation of CARP dimers, having been reported that the C-terminal AR fraction alone (105-319) cannot dimerize (Lun *et al.,* 2014). However, the current study shows that the N-terminally truncated variant, $CARP^{106-319}$, dimerizes as shown by SEC-MALLS data. The reason of this difference could be that the interaction of CARP AR region is of low affinity and poorly detectable by the BiYFP method used by Lun *et al.*, which is often affected by high background signal and is poorly quantitative (Lun *et al.,* 2014). Thermal denaturation data obtained both by CD and DSF showed that the thermal stability of $CARP^{106-319}$ is not high (Tm=43°C). In addition, the study of CARP and N2A complexation showed that complex formation causes the disruption of the dimer. Taken together, the relatively low thermal stability and the change of oligomeric state upon N2A complexation supports the deduction that the $CARP^{106-319}$ dimer formed by its AR domain is weak. It can be proposed that this C-terminal domain contributes additively to the N-terminus led dimerization of CARP.

4.1.2 The weakening of the CARP dimer by N-terminus removal is of biological relevance

SEC-MALLS measurements on complexed samples demonstrate that the CARP:N2A complex is formed by a 1:1 molar ratio. Considering that $CARP^{106-319}$ is a dimer and titin N2A a monomer, the interaction between CARP and titin N2A must require the breakage of the CARP dimer. This finding is supported by other evidence: Laure *et al* found that, *in vivo*, calpain 3 cleaves CARP, removing its N-terminus and that, as a consequence of the cleavage, the interaction of CARP with titin N2A is reinforced (Laure *et al.,* 2010). Our data suggest that the removal of the N-terminus of CARP weakens its dimer strength, providing a rational for how the cleavage enhances the CARP:titin N2A binding process.

4.13 The unique sequence element in titin N2A is structured

The titin UN2A element characteristic of cardiac isoforms of titin is widely believed to be an unstructured sequence that acts as an additional entropic spring in heart muscle. However, CD and NMR results from this study unambiguously showed that UN2A is structured, being over 60% rich in helical content. Thermal denaturation data obtained using CD indicated that this protein is very stable. During the course of this study, however, we could not obtain evidence that UN2A is a globular protein and the possibility remains of it being a highly anisometric, helical filamentous formation lacking a globular fold. This information will be sought in future research by applying the SAXS technique.

The knowledge of the UN2A structure obtained in this study changes the appreciation of its role in heart as a mechanical spring. It has been shown that the orientation of secondary structural elements relative to the applied force vector is a determinant of mechanical strength (Rohs *et al.,* 1999; Brockwell *et al.,* 2003). In rough terms, longitudinal shearing requires a breaking force approximately equivalent to the force needed to rupture one bond multiplied by the number of bonds to be broken, while the application of force orthogonally to secondary structure elements (peeling) loads each hydrogen bond in turn, bonds fail consecutively at a lower force and pass the load to the next bond (in a zipper fashion). Helices are, thereby, mechanically resistant secondary structure elements as their main chain hydrogen bonds align with the helical axis and, in turn, the helical axis aligns with the force vector during pulling. Thus, we predict the UN2A element to be mechanically stiffer than previously predicted and, thereby, to have a significant influence in heart mechanics. To test this hypothesis, we have established a collaboration with Prof. Henk Granzier (University of Arizona), an expert on titin mechanics. He will study the samples used in this work, both in isolation and in complex, using AFM to determine their mechanical properties.

4.2 New model of CARP:titin N2A interaction

4.21 The CARP binding site spans the UN2A-Ig81 domains

UN2A and Ig81 showed low binding affinity to CARP$^{106-319}$ as the complexes were not preserved during GF (*Fig.3.20; Fig.3.22*). However, the binding affinity between CARP$^{106-319}$ and UN2A-Ig81 was high enough that the complex remained formed during GF (*Fig.3.21*). Taken together, the CARP binding site spans both domains.

4.22 New interaction model: CARP regulates titin stiffness

The previously believed unstructured character of UN2A necessarily implied that the folded CARP protein, with a significantly shorter contour length than UN2A (roughly estimated as 35 nm for UN2A vs 5 nm for CARP's AR region), could only bind a short sequence in UN2A, covering only a small fraction of its total length *(Fig.4.1A)*. This model could not explain how CARP could modulate the stiffness of the N2A cardiac element in a notable way as to impact muscle physiology or affect more than a limited number of phosphorylation sites. However, our data now suggest that UN2A is folded (at least it has helical secondary structure) and that CARP binding also involves the subsequent Ig81 domain. Based on these findings, we propose a new model (*Fig.4.1B*) where CARP and UN2A have comparable contour lengths in their folded states, so that CARP binding now engages a large surface portion of UN2A. This model now explains how CARP binding could efficiently modulate N2A stiffness as well as more markedly influence its phosphorylation state. To provide evidence to support this model, the further structural characterization of UN2A in isolation (using SAXS and/or NMR) or as part of CARP complexes (crystallography) will be pursued in future work.

The new interaction model further explains that the CARP:titin N2A interaction may serve as both an immediate and a long-term mechanism to resist cardiomyocyte stretch overload. In chronic HF, the CARP:titin N2A interaction will work together with the switch of N2A:N2BA isoform, as this switch is considered to control the passive stiffness of heart muscle (Opitz *et al.,* 2004; Nagueh *et al.,* 2004; Makarenko *et al.,* 2004; Martin *et al.* 2007). However, compared with the

expression of CARP (90minutes, Miller *et al.,* 2003), changing the N2B:N2BA ratio is a relatively slow process (two weeks in canine DCM model, Opitz *et al.,* 2004). Therefore, the CARP:titin N2A interaction may serve as an immediate mechanism to strengthen the heart muscle and, thus, resist a sudden stretch overload.

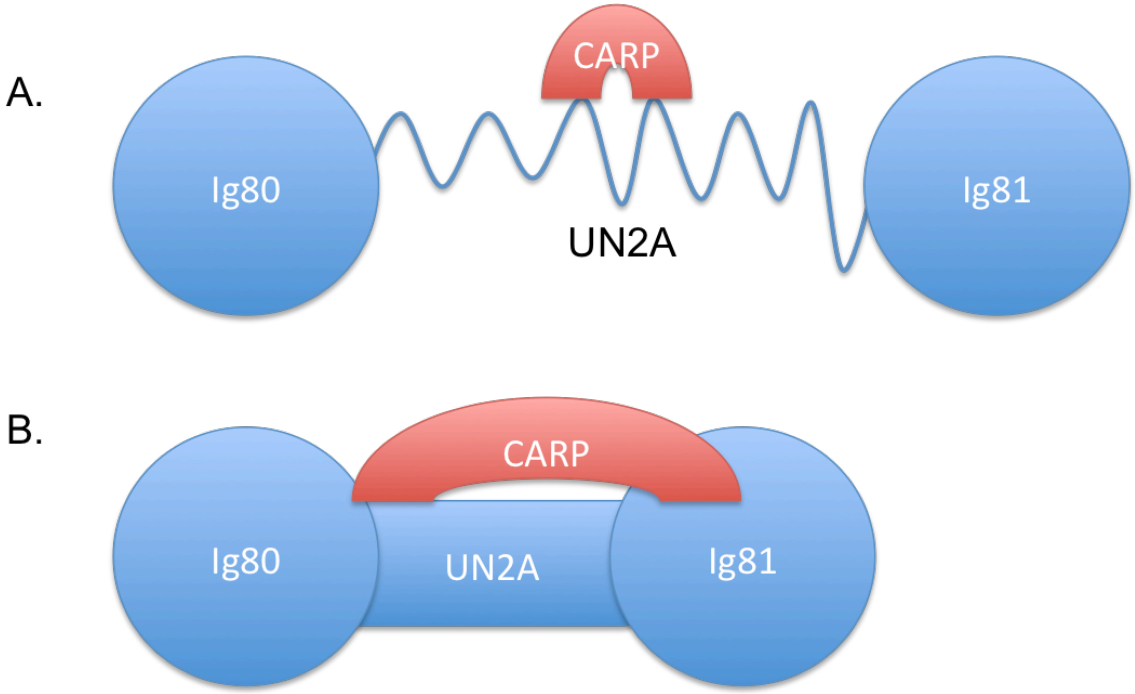


***Fig.4.1* CARP:titin N2A interaction models**

A) Original interaction model based on an unstructured UN2A where CARP bind only a small fraction of the titin scaffold.

B) New interaction model suggested by data gathered in this work in which CARP covers a larger UN2A portion as well as Ig81 making a more efficient regulator of titin stiffness and phosphorylation.

References

Andrade, M.A., Chacón, P., Merelo, J.J. and Morán, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. Prot. Engineering 6, 383-390.

Anurag M, Singh GP, Dash D. Location of disorder in coiled coil proteins is influenced by its biological role and subcellular localization: a GO-based study on human proteome. Mol Biosyst. 2012 Jan;8(1):346-52. doi: 10.1039/c1mb05210a. Epub 2011 Oct 25.

Arber S, Hunter JJ, Ross J Jr, Hongo M, Sansig G, Borg J, Perriard JC, Chien KR, Caroni P. MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. Cell. 1997 Feb 7;88(3):393-403.

Arimura T, Bos JM, Sato A, Kubo T, Okamoto H, et al. (2009) Cardiac ankyrin repeat protein gene (ANKRD1) mutations in hypertrophic cardiomyopathy. J Am Coll Cardiol 54: 334–342.

Bang ML, Mudry RE, McElhinny AS, Trombitás K, Geach AJ, Yamasaki R, Sorimachi H, Granzier H, Gregorio CC, Labeit S. Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple roles in Z-disc and I-band protein assemblies. J Cell Biol. 2001 Apr 16;153(2):413-27.

Barash IA, Bang ML, Mathew L, Greaser ML, Chen J, Lieber RL. Structural and regulatory roles of muscle ankyrin repeat protein family in skeletal muscle. Am J Physiol Cell Physiol. 2007 Jul;293(1):C218-27.

Baumeister A, Arber S, Caroni P (1997) Accumulation of muscle ankyrin repeat protein transcript reveals local activation of primary myotube end- compartments during muscle morphogenesis. J Cell Biol 139:1231–42

Bogomolovas J, Kathrin Brohm, Jelena Čelutkien, Giedr Balčiūnait, Daiva Bironait, Virginija Bukelskien, Dainius Daunoravičus, Christian C. Witt, Jens Fielitz, Virginija Grabauskien, Siegfried Labeit. Induction of Ankrd1 in dilated cardiomyopathy correlates with the heart failure progression. Biomed Res Int. 2015;2015:273936. doi: 10.1155/2015/273936. Epub 2015 Apr 16.

Breeden, L., and Nasmyth, K. (1987) Similarity between cell- cycle genes of budding yeast and fission yeast and the Notch gene of Drosophila, Nature 329, 651 - 4.

Brockwell DJ, Paci E, Zinober RC, Beddard GS, Olmsted PD, Smith DA, Perham RN, Radford SE. Pulling geometry defines the mechanical resistance of a beta-sheet protein. Nat Struct Biol 2003;10:731-7.

Byeon, I. J., Li, J., Ericson, K., Selby, T. L., Tevelev, A., Kim, H. J., O'Maille, P., and Tsai, M.-D. (1998) Tumor suppressor p16INK4A: Determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4, Mol. Cell 1, 421 - 31.

Cazorla O, Freiburg A, Helmes M, et al. Differential expression of cardiac titin isoforms and modulation of cellular stiffness. Circ Res 2000;86:59–67.

Compton, L.A. and Johnson, W.C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. Anal. Biochem. 155, 155-167.

Cheng WP, Wang BW, Lo HM, Shyu KG. Mechanical Stretch Induces Apoptosis Regulator TRB3 in Cultured Cardiomyocytes and Volume-Overloaded Heart. PLoS One. 2015 Apr 21;10(4):e0123235. doi: 10.1371/journal.pone.0123235. eCollection 2015.

Chu W, Burns DK, Swerlick RA, Presky DH (1995) Identification and characterization of a novel cytokine-inducible nuclear protein from human endothelial cells. J Biol Chem 270:10236–45

Duboscq-Bidot L, Charron P, Ruppert V, Fauchier L, Richter A, et al. (2009) Mutations in the ANKRD1 gene encoding CARP are responsible for human dilated cardiomyopathy. Eur Heart J 30: 2128–2136.

Eric H. Lee, Jen Hsin, Olga Mayans, Klaus Schulten Secondary and Tertiary Structure Elasticity of Titin Z1Z2 and a Titin Chain Model. Biophys J. 2007 Sep 1;93(5):1719-35. Epub 2007 May 11.

Ernesto Freire, Obdulio L. Mayorga, Martin Straume, Isothermal titration calorimetry, Anal. Chem., 1990, 62 (18), pp 950A–959A, DOI: 10.1021/ac00217a002

Granzier HL, Labeit S. The giant protein titin: a major player in myocardial mechanics, signaling, and disease. Circ Res. 2004 Feb 20;94(3):284-95.

Gregorio, C. C., Granzier, H., Sorimachi, H. & Labeit, S. 1999 Muscle assembly: a titanic achievement. Curr. Opin. Cell Biol. 11, 18 – 25.

Gaussin V, Tomlinson JE, Depre C, Engelhardt S, Antos CL, Takagi G, Hein L, Topper JN, Liggett SB, Olson EN, Lohse MJ, Vatner SF, Vatner DE. Common genomic response in different mouse models of beta-adrenergic-induced cardiomyopathy. Circulation. 2003 Dec 9;108(23):2926-33. Epub 2003 Nov 17.

Hanson J, Huxley HE. Structural basis of the cross-striations in muscle. Nature. 1953 Sep 19;172(4377):530-2.

Helmes M, Trombitás K, Centner T, Kellermayer M, Labeit S, Linke WA, Granzier H. Mechanically driven contour-length adjustment in rat cardiac titin's unique N2B sequence: titin is an adjustable spring. Circ Res. 1999 Jun 11;84(11):1339-52.

Henkel M, Redfield M, Weston SA, Gerber Y, Roger VL. Death in heart failure. A community perspective. Circ Heart Fail. 2008;1:91-7.

Hidalgo C, Granzier H. Tuning the molecular giant titin through phosphorylation: role in health and disease. Trends Cardiovasc Med. 2013 Jul;23(5):165-71. doi: 10.1016/j.tcm.2012.10.005. Epub 2013 Jan 5.

Howarth JW, Ramisetti S, Nolan K, Sadayappan S, Rosevear PR. Structural insight into unique cardiac myosin-binding protein-C motif: a partially folded domain. J Biol Chem. 2012 Mar 9;287(11):8254-62. doi: 10.1074/jbc.M111.309591. Epub 2012 Jan 10.

Ihara Y, Suzuki YJ, Kitta K, Jones LR, Ikeda T. Modulation of gene expression in transgenic mouse hearts overexpressing calsequestrin. Cell Calcium. 2002 Jul;32(1):21-9.

Ishiguro N, Baba T, Ishida T, Takeuchi K, Osaki M, Araki N, Okada E, Takahashi S, Saito M, Watanabe M, Nakada C, Tsukamoto Y, Sato K, Ito K, Fukayama M, Mori S, Ito H, Moriyama M. Carp, a cardiac ankyrin-repeated protein, and its new homologue, Arpp, are differentially expressed in heart, skeletal muscle, and rhabdomyosarcomas. Am J Pathol. 2002 May;160(5):1767-78.

Jeyaseelan R, Poizat C, Baker RK, Abdishoo S, Isterabadi LB, Lyons GE et al. (1997) A novel cardiac-restricted target for doxorubicin. CARP, a nuclear modulator of gene expression in cardiac progenitor cells and cardiomyocytes. J Biol Chem 272:22800–8

Johnson, W. C., Jr. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. Proteins: Struct. Funct. Genet. 35, 307–312.

J Yang, R Yan, A Roy, D Xu, J Poisson, Y Zhang. The I-TASSER Suite: Protein structure and function prediction. Nature Methods, 12: 7-8, 2015.

Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010 Feb;66(Pt 2):125-32. doi: 10.1107/S0907444909047337. Epub 2010 Jan 22.

Kang PM, Izumo S. Apoptosis and heart failure: A critical review of the literature. Circ Res. 2000 Jun 9;86(11):1107-13.

Kemp, T. J., Sadusky, T. J., Saltisi, F., Carey, N., Moss, J., Yang, S. Y. et al. (2000). Identification of Ankrd2, a novel skeletal muscle gene coding for a stretch-responsive ankyrin-repeat protein. Genomics, 66, 229–241.

Krüger M, Linke WA. Protein kinase-A phosphorylates titin in human heart muscle and reduces myofibrillar passive tension. J Muscle Res Cell Motil. 2006;27(5-7):435-44. Epub 2006 Aug 4.

Krüger M, Kötter S, Grützner A, Lang P, Andresen C, Redfield MM, Butt E, dos Remedios CG, Linke WA. Protein kinase G modulates human myocardial passive

stiffness by phosphorylation of the titin springs. Circ Res. 2009 Jan 2;104(1):87-94. doi: 10.1161/CIRCRESAHA.108.184408.

Labeit S, Barlow DP, Gautel M, et al. A regular pattern of two types of 100- residue motif in the sequence of titin. Nature 1990;345:273–6.

Labeit S, Kolmerer B. Titins: giant proteins in charge of muscle ultrastructure and elasticity. Science. 1995 Oct 13;270(5234):293-6.

Laure L, Danièle N, Suel L, Marchand S, Aubert S, Bourg N, Roudaut C, Duguez S, Bartoli M, Richard I. A new pathway encompassing calpain 3 and its newly identified substrate cardiac ankyrin repeat protein is involved in the regulation of the nuclear factor-κB pathway in skeletal muscle. FEBS J. 2010 Oct;277(20):4322-37. doi: 10.1111/j.1742-4658.2010.07820.x. Epub 2010 Sep 22.

Laustsen PG, Russell SJ, Cui L, Entingh-Pearsall A, Holzenberger M, Liao R, Kahn CR. Essential role of insulin and insulin-like growth factor 1 receptor signaling in cardiac development and function. Mol Cell Biol. 2007 Mar;27(5):1649-64. Epub 2006 Dec 22.

LeWinter MM, Granzier H. Cardiac titin: a multifunctional giant. Circulation. 2010 May 18;121(19):2137-45. doi: 10.1161/CIRCULATIONAHA.109.860171.

Li, X., P. Romero, M. Rani, A.K. Dunker, and Z. Obradovic, Genome Informatics, 1999, 10:30-40.

Lun AS, Chen J, Lange S. (2014) Probing muscle ankyrin-repeat protein (MARP) structure and function, Anat Rec (Hoboken). 2014 Sep;297(9):1615-29. doi: 10.1002/ar.22968.

Maeda T, Sepulveda J, Chen HH, Stewart AF (2002) Alpha(1)-adrenergic activation of the cardiac ankyrin repeat protein gene in cardiac myocytes. Gene 297:1–9

Makarenko I, Opitz CA, Leake MC, Neagoe C, Kulke M, Gwathmey JK, del Monte F, Hajjar RJ, Linke WA. Passive stiffness changes caused by upregulation of compliant titin isoforms in human dilated cardiomyopathy hearts. Circ Res. 2004 Oct 1;95(7):708-16. Epub 2004 Sep 2.

Martin M. LeWinter , Yiming Wu, Siegfried Labeit, Henk Granzier (2007) Cardiac titin: Structure, functions and role in disease Clinica Chimica Acta 375 (2007) 1–9

Mikhailov AT, Torrado M. The enigmatic role of the ankyrin repeat domain 1 gene in heart development and disease. Int J Dev Biol 2008;52: 811 – 821.

Miller MK, Bang ML, Witt CC, Labeit D, Trombitas C, et al. (2003) The muscle ankyrin repeat proteins: CARP, ankrd2/Arpp and DARP as a family of titin filament-based stress response molecules. J Mol Biol 333: 951–964

Miller MK, Granzier H, Ehler E, Gregorio CC. The sensitive giant: the role of titin-based stretch sensing complexes in the heart. Trends Cell Biol. 2004 Mar;14(3):119-26.

Moulik M, Vatta M, Witt SH, Arola AM, Murphy RT, et al. (2009) ANKRD1, the gene encoding cardiac ankyrin repeat protein, is a novel dilated cardiomyopathy gene. J Am Coll Cardiol 54: 325–333.

Muller-Seitz M, Kaupmann K, Labeit S, Jockusch H. Chromosomal localization of the mouse titin gene and its relation to "muscular dystrophy with myositis" and nebulin genes on chromosome 2. Genomics 1993;18: 559–61.

Nagueh SF, Shah G, Wu Y, Torre-Amione G, King NM, et al. (2004) Altered titin expression, myocardial stiffness, and left ventricular function in patients with dilated cardiomyopathy. Circulation 110: 155–162.

National Clinical Guideline Centre (UK). Chronic Heart Failure: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care: Partial Update.

Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc. 2007;2(9):2212-21.

Opitz CA, Leake MC, Makarenko I, Benes V, Linke WA. Developmentally regulated switching of titin size alters myofibrillar stiffness in the perinatal heart. Circ Res. 2004 Apr 16;94(7):967-75. Epub 2004 Feb 26.

Rohs R, Etchebest C, Lavery R. Unraveling proteins: a molecular mechanics study. Biophys J 1999; 76:2760-8.

Rossi E, Faiella A, Zeviani M, et al. Order of six loci at 2q24–q31 and orientation of the HOXD locus. Genomics 1994;241:34–40.

Shi Y, Reitmaier B, Regenbogen J, Slowey RM, Opalenik SR, Wolf E, Goppelt A, Davidson JM. CARP, a cardiac ankyrin repeat protein, is up-regulated during wound healing and induces angiogenesis inexperimental granulation tissue. Am J Pathol. 2005 Jan;166(1):303-12.

Squire JM. Architecture and function in the muscle sarcomere. Curr Opin Struct Biol. 1997 Apr;7(2):247-57.

Sreerema, N., Venyaminov, S.Y., and Woody, R.W. (1999) Estimation of the number of helical and strand segments in proteins using CD spectroscopy. Protein Sci. 8, 370-380.

Taylor M, Graw S, Sinagra G, Barnes C, Slavov D, Brun F, Pinamonti B, Salcedo EE, Sauer W, Pyxaras S, Anderson B, Simon B, Bogomolovas J, Labeit S, Granzier H, Mestroni L. Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes. Circulation. 2011 Aug 23;124(8):876-85. doi: 10.1161/CIRCULATIONAHA.110.005405. Epub 2011 Aug 1.

Tevelev, A., Byeon, I. J., Selby, T., Ericson, K., Kim, H. J., Kraynov, V., and Tsai, M.-D. (1996) Tumor suppressor p16INK4A: Structural characterization of wild-type and mutant proteins by NMR and circular dichroism, Biochemistry 35, 9475 - 87.

Torrado M, López E, Centeno A, Castro-Beiras A, Mikhailov AT. Left-right asymmetric ventricular expression of CARP in the piglet heart: regional response to experimental heart failure. Eur J Heart Fail. 2004 Mar 1;6(2):161-72.

Torrado, M., Nespereira, B. and Mikhailov, A.T. (2005b). Fetal cardiac control genes: implications for postnatal heart growth and heart disease. Trends Dev Biol 1: 27-38.

Van Stokkum, I.H.M., Spoelder, H.J.W., Bloemendal, M., Van Grondelle, R., and Groen, F.C.A. (1990) Estimation of protein secondary structure and error analysis from CD spectra. Anal. Biochem. 191, 110-118.

Wallace and Janes, Modern Techniques in Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy, 2009. ISBN: 978-1-60750-000-1

Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics. 2014 Sep 8;47:5.6.1-5.6.32. doi: 10.1002/0471250953.bi0 506s47.

Wei YJ, Cui CJ, Huang YX, Zhang XL, Zhang H, et al. (2009) Upregulated expression of cardiac ankyrin repeat protein in human failing hearts due to arrhythmogenic right ventricular cardiomyopathy. Eur J Heart Fail 11: 559– 566.

Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W668-73.

Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers. 2008 May;89(5):392-400.

Wienken CJ, Baaske P, Rothbauer U, Braun D, Duhr S. Protein-binding assays in biological liquids using microscale thermophoresis. Nat Commun. 2010 Oct 19;1:100. doi: 10.1038/ncomms1093.

Witt SH, Labeit D, Granzier H, Labeit S, Witt CC. 2005. Dimerization of the cardiac ankyrin protein CARP: implications for MARP titin-based signaling. J Muscle Res Cell Motil 26:401–408.

Xue, B., R. L. DunBrack, R.W. Williams, A.K. Dunker, and V. N. Uversky (2010) "PONDR-Fit: A meta-predictor of intrinsically disordered amino acids," Biochim. Biophys. Acta (in press) doi:10.1016/j.bbapap.2010.01.011

Y. Zhang, D. Xu, Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins, 2012, 80, 1715-1735

Zolk O, Frohme M, Maurer A, Kluxen FW, Hentsch B, et al. (2002) Cardiac ankyrin repeat protein, a negative regulator of cardiac gene expression, is augmented in human heart failure. Biochem Biophys Res Commun 293: 1377– 1382.

Zolk O, Marx M, Jäckel E, El-Armouche A, Eschenhagen T. Beta-adrenergic stimulation induces cardiac ankyrin repeat protein expression: involvement of protein kinase A and calmodulin-dependent kinase. Cardiovasc Res. 2003 Sep 1;59(3):563-72.